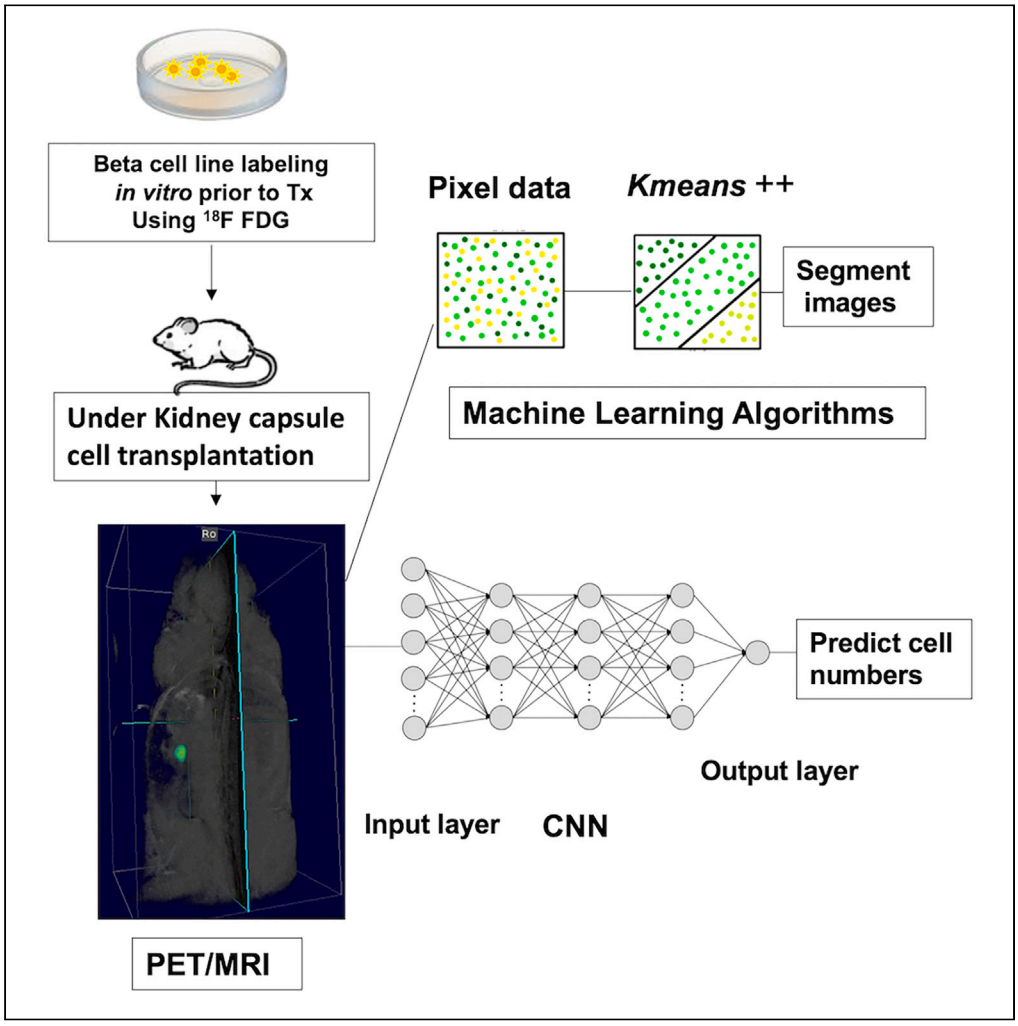
iScience

CellPress
OPEN ACCESS

Article

Deep learning-enabled quantification of simultaneous PET/MRI for cell transplantation monitoring



Hasaan Hayat, Rui Wang, Aixia Sun, ..., Guo-Wei Wei, Jinda Fan, Ping Wang

fanjinda@msu.edu (J.F.)
wangpin4@msu.edu (P.W.)

Highlights

Algorithms were developed for the analysis of PET/MRI of transplanted cell

Three algorithms were trained and tested with *in vitro* and *in vivo* datasets

3D CNN predicts actual transplanted cell numbers in cell-transplanted mouse models

We provide a novel tool for monitoring islet transplantation through PET/MRI

Hayat et al., iScience 26, 107083
July 21, 2023
https://doi.org/10.1016/j.isci.2023.107083

iScience

CellPress
OPEN ACCESS

Article

Deep learning-enabled quantification of simultaneous PET/MRI for cell transplantation monitoring

Hasaan Hayat,[1,2,3] Rui Wang,[4] Aixia Sun,[1,2] Christiane L. Mallett,[2,5] Saumya Nigam,[1,2] Nathan Redman,[5,6] Demarcus Bunn,[5,6] Elvira Gjelaj,[1,7] Nazanin Talebloo,[1,8] Adam Alessio,[2,5,6,9] Anna Moore,[1,2] Kurt Zinn,[2,5] Guo-Wei Wei,[4,9,10] Jinda Fan,[2,5,8,*] and Ping Wang[1,2,11,*]

SUMMARY

Current methods of *in vivo* imaging islet cell transplants for diabetes using magnetic resonance imaging (MRI) are limited by their low sensitivity. Simultaneous positron emission tomography (PET)/MRI has greater sensitivity and ability to visualize cell metabolism. However, this dual-modality tool currently faces two major challenges for monitoring cells. Primarily, the dynamic conditions of PET such as signal decay and spatiotemporal change in radioactivity prevent accurate quantification of the transplanted cell number. In addition, selection bias from different radiologists renders human error in segmentation. This calls for the development of artificial intelligence algorithms for the automated analysis of PET/MRI of cell transplantations. Here, we combined *K*-means++ for segmentation with a convolutional neural network to predict radioactivity in cell-transplanted mouse models. This study provides a tool combining machine learning with a deep learning algorithm for monitoring islet cell transplantation through PET/MRI. It also unlocks a dynamic approach to automated segmentation and quantification of radioactivity in PET/MRI.

INTRODUCTION

The combination of positron emission tomography (PET) and magnetic resonance imaging (MRI) has been a tremendous milestone in multi-modal imaging. Simultaneous PET/MRI imaging is a powerful tool for both functional and structural imaging of a biological target. This imaging modality has synergistic advantages over alternative imaging modalities such as either PET or MRI alone as it provides two-pronged validation of the imaging target and provides insight on cell metabolism.[1,2] PET/MRI also allows better image mapping and improved resolution over PET alone, allowing for improved visualization of anatomical structures.[3–5] Similarly, MRI alone provides minimal sensitivity and specificity of the functional information in comparison to PET.[6] Simultaneous PET/MRI also holds an advantage over PET/computed tomography (CT) since MRI provides excellent soft-tissue contrast and lacks the exposure of ionizing radiation; therefore patients are not exposed to its harmful effects and can undergo multiple scans.[5,7–10] PET/MRI utilizes the advantages of both techniques, making it a formidable candidate for cellular therapy imaging.

There have been new radiolabeling techniques used on cell transplantations which assist with molecular tracking using PET/MRI.[11–13] One of those techniques consists of the synthesis of ^{64}Cu-CD45 and ^{89}Zr-CD45 performed by Tarantal et al. where PET was then used for the *in vivo* tracking of the radiolabeled cells.[11] Radiolabeled cell transplantations have been used for targeting surface antigens for leukemias, Hodgkin's lymphomas, and other hematologic malignancies.[14,15] In cell-based therapy, the route of administration plays a critical role in the delivery of the therapeutic cells. To avoid repeated injections (traditional route of administration) which risk embolism, PET allows visualization of molecular and cellular alterations and monitoring of the stem cells.[16] These advancements in medical imaging technologies are revolutionizing the way physicians monitor treatment. However, although simultaneous PET/MRI is expanding its scope as an imaging modality with functional and anatomical flexibility, researchers have not yet developed a standardized method for image segmentation and subsequent analysis of the complex data.[17]

[1]Precision Health Program, Michigan State University, 766 Service Road, Rm. 2020, East Lansing, MI 48823, USA

[2]Department of Radiology, College of Human Medicine, Michigan State University, East Lansing, MI, USA

[3]College of Human Medicine, Michigan State University, East Lansing, MI, USA

[4]Department of Mathematics, College of Natural Science, Michigan State University, East Lansing, MI, USA

[5]Institute for Quantitative Health Science and Engineering, Department of Biomedical Engineering, Michigan State University, East Lansing, MI, USA

[6]Department of Biomedical Engineering, College of Engineering, Michigan State University, East Lansing, MI, USA

[7]Lyman Briggs College, Michigan State University, East Lansing, MI, USA

[8]Department of Chemistry, College of Natural Science, Michigan State University, East Lansing, MI, USA

[9]Departments of Computational Mathematics, Science, and Engineering (CMSE), College of Natural Science, Michigan State University, East Lansing, MI, USA

[10]Department of Electrical and Computer Engineering, College of Engineering, Michigan State University, East Lansing, MI, USA

[11]Lead contact

*Correspondence: fanjinda@msu.edu (J.F.), wangpin4@msu.edu (P.W.)

https://doi.org/10.1016/j.isci.2023.107083

More importantly, the lack of computational methods to assist in dynamic analysis of PET/MRI data limits the reliability and clinical translatability of novel techniques and therapies.

The expanding need for artificial intelligence (AI), and particularly, deep learning (DL), is becoming highly evident in the context of image segmentation and quantification for analysis and monitoring of treatments and therapies, especially for newer modalities such as PET/MRI which present data with spatiotemporal heterogeneity. This heterogeneity limits the ability to definitively analyze the image data with high reliability and link the analysis to a quantitative metric of interest or a diagnosis. Currently, AI algorithms are prominently used for the diagnosis of diseases especially in medical specialties of oncology for assessment of tumors and neurology for prediction of Alzheimer's and classification of brain lesions through medical image segmentation and analysis.[18,19] AI, including machine leaning (ML) and DL, can also be trained by datasets and makes accurate predictions as the response to treatment or the mortality after the cell transplantation.[20,21]

The role of PET in monitoring cell transplantation coupled with the potential use of simultaneous PET/MRI for greater spatial resolution is particularly important for treating diseases such as type 1 diabetes (T1D), where islet cell transplantation therapy requires extensive imaging for detection of graft loss and adequate intervention.[22,23] However, due to variability among PET/MRI images in terms of signal distribution, signal size, indeterminate boundaries of true signal, and the interrater variability and human bias in selection of a region of interest (ROI) from simultaneous PET/MRI images, this task is increasingly challenging and unreliable. Furthermore, the time-dependent decay of signal over time in PET imaging and spatiotemporal change in cellular signal distribution from PET/MRI create additional difficulty and add a degree of uncertainty to estimating metrics that depend on radioactivity from the scan. Such metrics include quantification of injected dose, calculation of standard uptake value (SUV) in target areas, and estimation of transplanted cell number. Additionally, the true transplanted cell number is often unknown without the use of extensive measures and manual measurement of dose injected and the corresponding radioactivity, and it is difficult to correlate this to cell number without referring to a previously determined standard after mathematical transformation of the dose reading at the time of scanning. This calls for the development of an AI algorithm for standardized, automated segmentation and quantification of such metrics from simultaneous PET/MRI scans for longitudinal monitoring and assessment of cell transplantations.

To address this problem, we aim to develop a canonical AI algorithm which uniquely combines unsupervised ML and DL for the purposes of PET/MRI segmentation and quantification, respectively, for analysis of transplanted cells labeled with ^{18}F-fluorodeoxyglucose (FDG) *in vivo*. Unsupervised ML is a robust approach to implementing AI algorithms in the field of biomedical imaging. Unlike supervised ML which requires a large volume of training data to train the deep neural network and is often employed for quantification tasks, this method of ML focuses on initial parameter optimization to perform on a wide domain of data for segmentation tasks. This requires less volume of data for initial construction and optimization of the algorithm and does not need the level of human interference required in supervised algorithms such as DL. This is especially useful when deployed for image segmentation because large volumes of manually labeled images are not necessary for optimal segmentation performance of the algorithm. For standardized tasks such as selection of an ROI for image segmentation, unsupervised ML has advantages over supervised ML because it is independent of the shape and distribution of signal in an ROI.[24] Therefore, to address selection bias and interrater variability in evaluation of an ROI from simultaneous PET/MRI imaging, we aim to use *K-means++*, a clustering algorithm in the domain of unsupervised ML, for segmentation of the desired ROI from a simultaneous PET/MRI scan (Figure 1). In our recent published studies, we have demonstrated the capability of our *K*-means++ algorithm to segment the sufficient ROI from 2D and 3D magnetic particle imaging (MPI) of human islets and stem cell transplantations *in vivo*.[24,25] The similar, positive-contrast nature of PET images encouraged us to employ this algorithm for segmentation of the PET ROIs in our scans. For the more challenging task of analysis of the PET signal from the 3D images and estimation of cell number that was truly transplanted, we aim to use a 3D convolutional neural network (CNN), which is a conventional DL algorithm, for prediction of the activity within the ROI in the sample (and mice) at the time of scanning and classification of the estimated cell number in the ROI based on this value. The core functionality of the 3D CNN will effectively augment the role of the conventional dose calibrator that is used to measure activity before and after scanning of the sample/subject, with the aim of automating and standardizing this process and using the information retrieved to make intended transformations, classifications, and predictions (Figure 1). The goal of the study is to establish the effectiveness of a combined

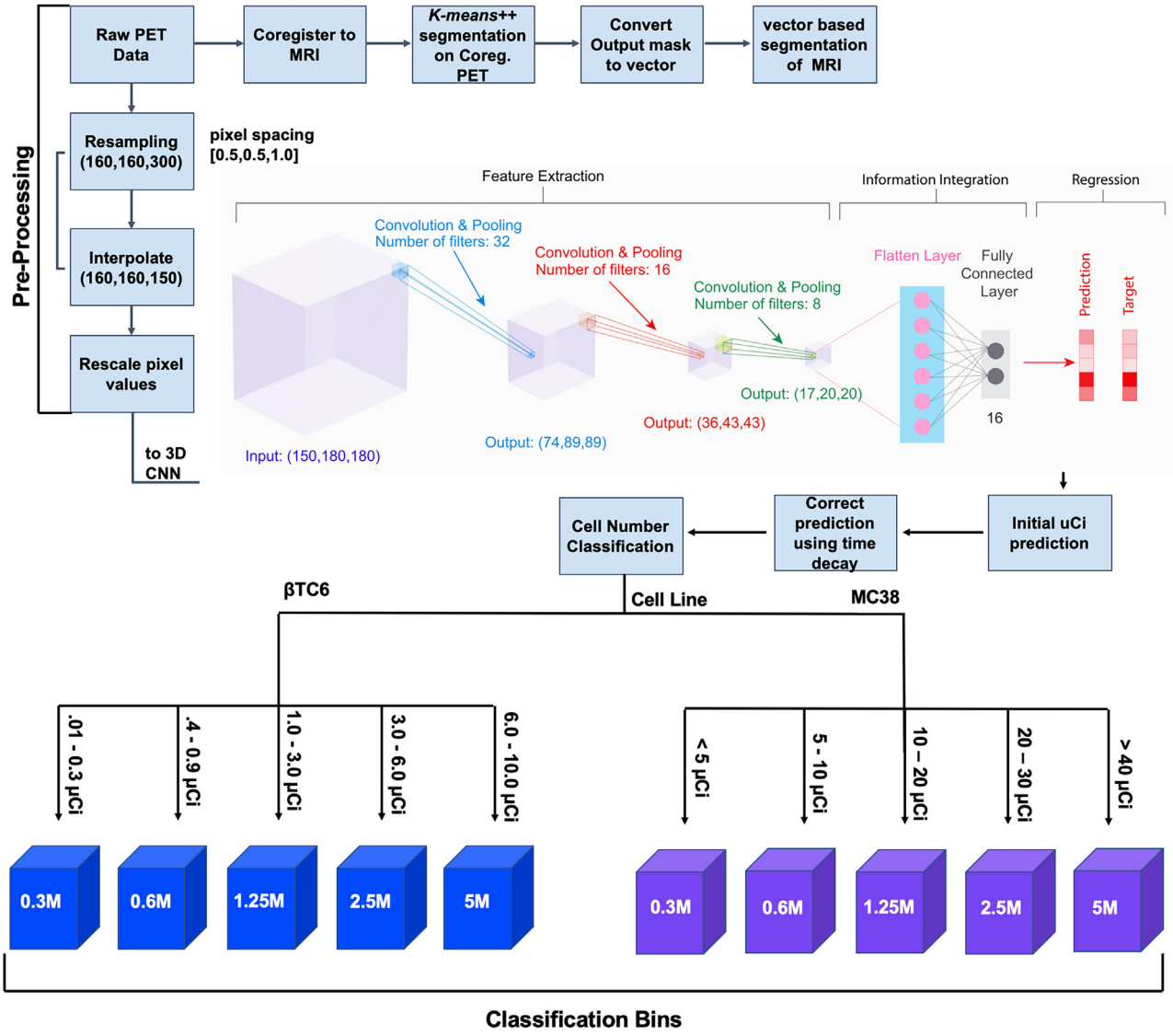


Figure 1. Schematic overview of the combined *K*-means++ segmentation and deep learning algorithm for segmentation of the co-registered PET/MRI data and simultaneous prediction of radioactivity and cell number within a 3D PET image sequence

For the 3D CNN model, as illustrated in the diagram, four convolutional layers and four average pooling layers are included in the 3D CNN model. Here, the kernel size of each convolutional layer is set to be three and the pooling size is two. Output predictions from the 3D CNN are transformed into new radioactivity values by the time-decay equation and undergo subsequent sorting into one of the cell line-specific classification bins to estimate cell number based on transformed final radioactivity value.

unsupervised ML and DL algorithm for both segmentation and analysis of simultaneous PET/MRI images for prediction of the transplanted cell number *in vivo* (Figures 1 and 2). If proven viable, this combined algorithm can be applied to numerous other applications of segmentation and analysis of simultaneous PET/MRI such as monitoring and quantifying injected radiolabeled nanodrug, assessing tumor response to chemotherapy, and diagnostic imaging with radiolabeled probes.

RESULTS

K-means++ segmentation

Initial testing of the *K*-means++ clustering algorithm on the raw PET Digital Imaging and Communications in Medicine (DICOM) images using the k values of 2, 3, and 4 for an arbitrary cluster of ^{18}F-FDG-labeled cells in agarose gel (Figure 3A) and ^{18}F-FDG-labeled cells in an "M"-shaped agarose gel phantom (Figure 3B)

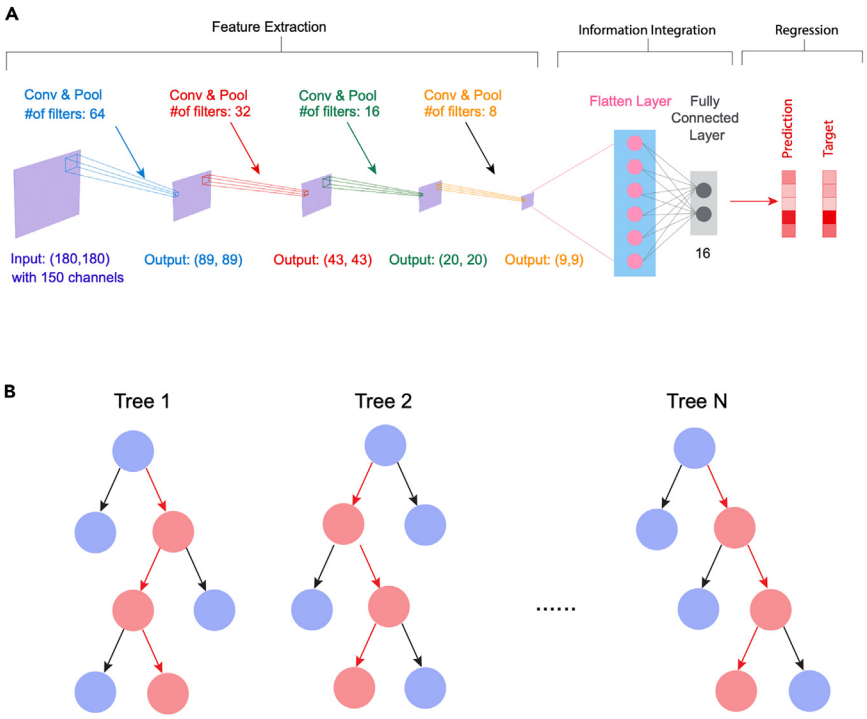


Figure 2. Inlustraions of 2D CNN and GBDT models

(A) Schematic overview of the 2D CNN model. Four convolutional layers and four max pooling layers are included in the 2D CNN model. Here, the kernel size of each convolutional layer is set to be three and the pooling size is two.

(B) Illustration of GBDT model with *N* gradient boosting trees.

shows differential segmentation contingent upon different k values. As indicated by the increased segmentation of true pixels when a k value of 4 is selected for clustering segmentation, the algorithm overshoots its clustering prediction and discards pixel values that pertain to the labeled cells in the phantom in both the arbitrary and "M"-shaped phantoms. Contrarily, when a k value of 2 is selected for clustering-based segmentation, the algorithm includes a great deal of noise pollution or signal that does not pertain to cells within the phantom. With a k value of 3 it is evident that only signal from the actual region of labeled cells within the phantom is clustered and segmented, and a reliable segmentation of the PET ROI can be extracted. To further test this, the elbow method was employed for testing of the minimum sum squared error (MSSE) among different k values. From this analysis, an optimal k value of 3 was determined for clustering segmentation and thus used for segmentation of the ROI throughout the rest of the study.

CNN training and comparative analysis with alternative algorithms

The comparative analysis of the 3D CNN with both 2D CNN and gradient boosting decision tree (GBDT) models revealed superior performance of the 3D CNN over the alternative algorithms used in the study (Figure 4). The 3D CNN had the lowest root mean square error (RMSE) of 4.2 $\pm$ 0.3 and mean absolute error (MAE) of 2.6 $\pm$ 0.1 as compared to the 2D CNN and GBDT ($p < 0.05$) (Figure 4A). It also maintained the highest Pearson correlation coefficient (PCC) value of 97% as compared to the other two algorithms (Figure 4B). The GBDT performed with a lower RMSE of 6.1 (+/− 0.3) than the 2D CNN (7.2 $\pm$ 0.2) with similar PCC values between the two algorithms—95% for the GBDT and 92% for 2D CNN. However, the 2D CNN did have a lower MAE (3.6 $\pm$ 0.2) than GBDT (4.0 $\pm$ 0.1) which is indicative of better training and overall algorithm performance in comparison to ground truth for the 2D CNN when compared to the GBDT algorithm ($p < 0.05$) (Figure 4A). However, the 3D CNN maintained the most optimal values across all 3 categories of performance measures. The 3D CNN and 2D CNN were trained on two datasets, dataset 1 and dataset 2, and their decreasing loss throughout ten iterations of training was evaluated for comparison among the models (Figure 4C). Further analysis on the GBDT was omitted from the study due to the model's inferior performance in comparison to the CNNs as indicated by its high MAE. The respective

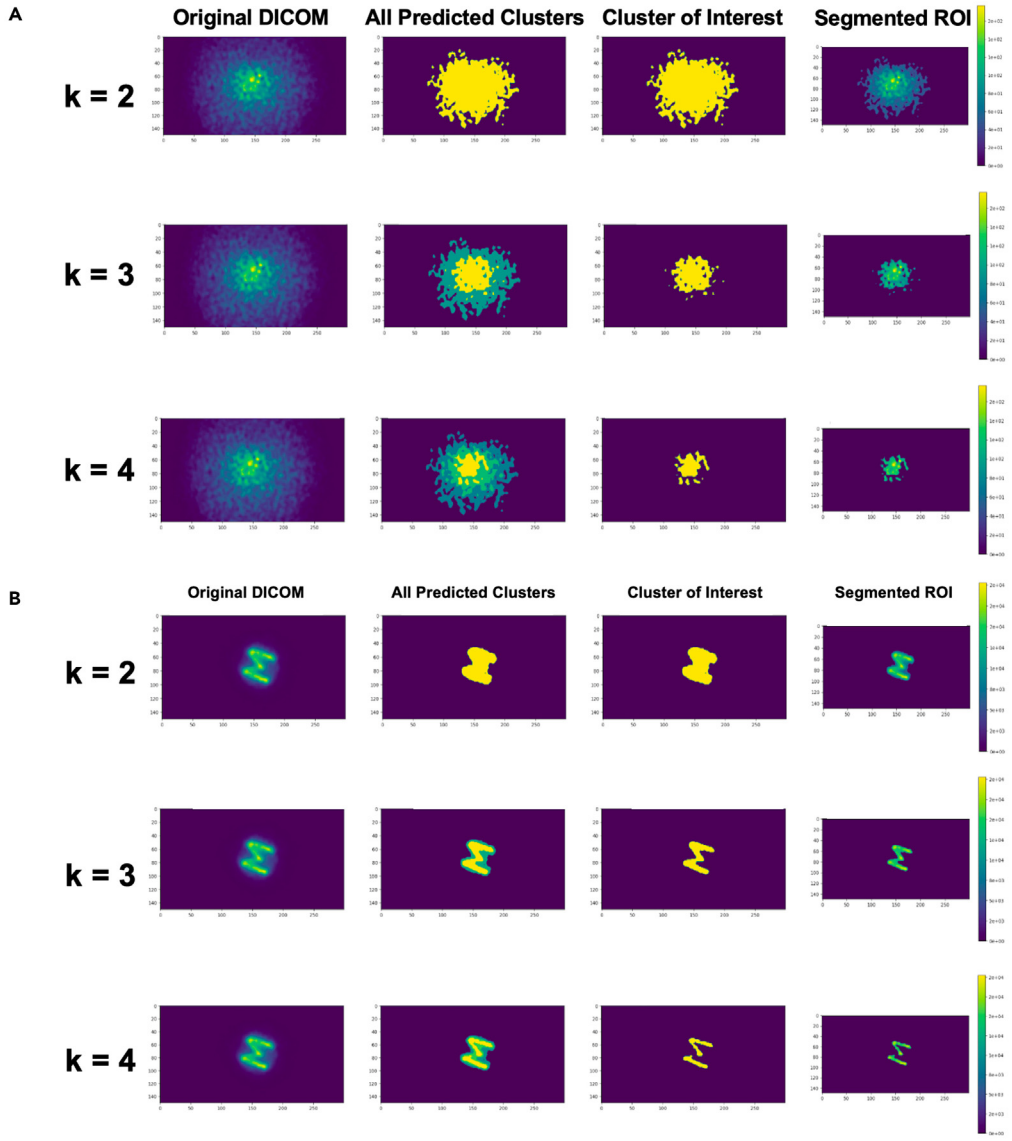


Figure 3. Initial results from *K*-means++ segmentation of PET imaging of circular and "M"-shaped phantoms with K values of 2, 3, and 4

(Left to right; Original DICOM, All predicted clusters, Cluster of interest, and segmented ROI using cluster of interest).

(A) Indicates segmentation results from circle phantom segmentation with K values of 2, 3, and 4.

(B) Indicates segmentation results from "M"-shaped phantom segmentation for K values of 2, 3, and 4.

training and validation loss curves over 10 epochs of training for the 3D CNN and the 2D CNN are depicted in Figures S1A and S1B, respectively. The first dataset, dataset 1, consisted of approximately 102 full 3D PET volumes (n = 102). A large majority of the data volumes in dataset 1 consisted of images that corresponded to radioactivity values greater than 10 μCi. Dataset 2 had a total of 150 (n = 150) full 3D PET image volumes, consisting of all the images in dataset 1 with additional images of radioactivity values less than 10 μCi. Dataset 2 also consisted of images with non-uniform pixel configuration, containing disjointed and scattered pixel matrices instead of the largely uniform ROIs present in the first dataset. This permitted dynamic training of the algorithm to predict future ROIs of unknown radioactivity despite a non-continuous, segmented spatial configuration of pixels corresponding to PET signal in the images. The 3D CNN maintained the lowest loss value upon training on both datasets for radioactivity prediction (Figure 4C). An improvement in 3D CNN algorithm performance is seen between training on dataset 1 and dataset 2, as indicated by the overall decreased loss value at the end of training iteration ten. A slightly lower loss value

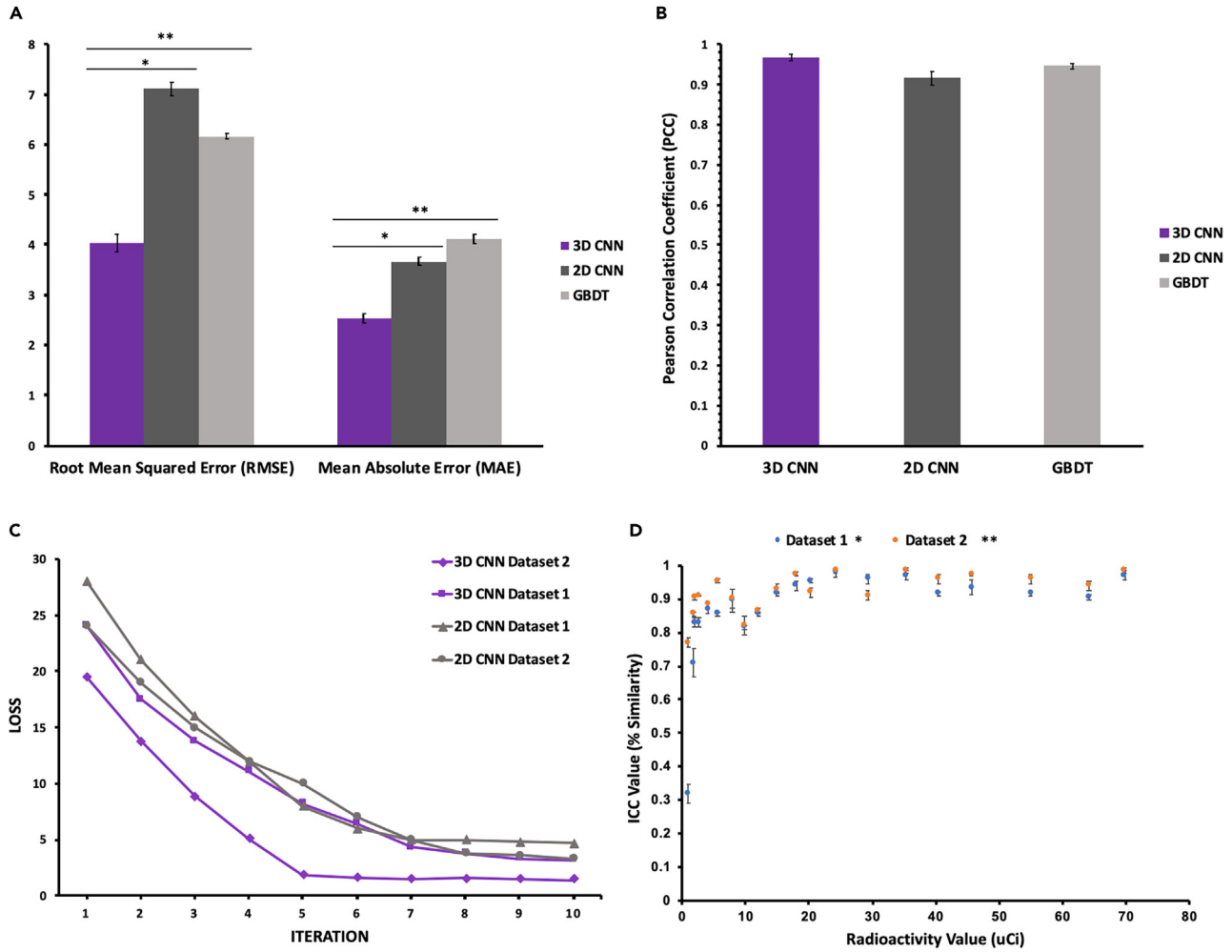


Figure 4. Comparison, testing, and training of 3D CNN

(A) Bar graph showing comparison of predictive performance of the 3D CNN, 2D CNN, and GBDT algorithms via analysis of root mean squared error (RMSE) and mean absolute error (MAE) (*$p < 0.05$, **$p < 0.05$, the Student's t test).

(B) Bar graph showing comparison of Pearson correlation coefficient (PCC) values between 3D CNN, 2D CNN, and GBDT algorithms.

(C) Loss graph of ten training rounds (iterations) comparing the 3D CNN and 2D CNN for both dataset 1 and dataset 2.

(D) Intraclass correlation coefficient analysis of 3D CNN's predictive accuracy on a range of images with 0 to 70μCi radioactivity (*$p = 0.001$, **$p = 0.001$, the F test).

is seen for both dataset 1 and dataset 2 training of the 2D CNN as compared to the 3D CNN (Figure 4C). Despite the relative similar PCC values between the three algorithms, the significantly lower RMSE, MAE, and loss values for training of the 3D CNN provided enough support to elect this model as the DL algorithm of choice for use throughout the rest of the study for radioactivity prediction from 3D PET image sequences and subsequent cell number classification.

To assess the predictive accuracy of the 3D CNN, radioactivity prediction values of PET image sequences were compared to true dose calibrator values by the intraclass correlation coefficient (ICC) analysis (Figure 4D). ICC analysis was performed for the two datasets trained on the algorithm in sequential fashion. As mentioned previously, dataset 1 consisted of image volumes corresponding to radioactivity values greater than 10 μCi. Despite the inherent limitation of dataset 1, ICC values for all radioactivity values in the range of 0–70 μCi maintained strong agreement with the dose calibrator as indicated by the ICC values of 0.70–0.90 (70–90% agreement) for images with radioactivity values in the range of 0–10 μCi and greater than 0.90 (90%) agreement for radioactivity values beyond 10 μCi. Upon ICC analysis of dataset 2, which contained images reflecting a broader range of radioactivity values especially those of lower μCi,

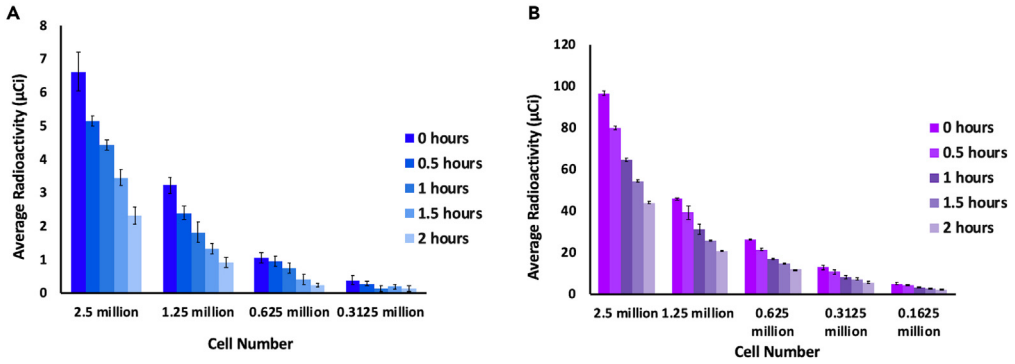


Figure 5. Characterization of radioactive uptake and establishment of standardized classification bins for different cell numbers in Beta-TC6 and MC38 cell lines labeled with 18F-FDG

(A) Characterization of average radioactivity (n = 3) of 2.5M, 1.25M, 0.625M, and 0.3125M MC-38 cells measured at time intervals 0.0H, 0.5H, 1.0H, 1.5H, 2.0H (M = million; H = hours).

(B) Characterization of average radioactivity (n = 3) of 2.5 M, 1.25 M, 0.625 M, and 0.3125 M Beta-TC6 cells measured at time intervals 0.0H, 0.5H, 1.0H, 1.5H, 2.0H.

agreement values markedly increased from prior values to greater than 90% agreement for radioactivity values lower than 10 μCi, indicating that the neural network trained with greater accuracy in predicting on this lower range of radioactivity. After training on dataset 2, more than 95% of the range of radioactivity predictions had greater than 90% agreement with the values read by the dose calibrator. The most significant increases in agreement were observed for values less than 10 μCi as the sample size of PET image volumes in this range of radioactivity was deliberately increased in the second dataset to improve predictions.

Cell number classification bins based on radioactive dose uptake

Dose calibrator-based readings of radioactivity were taken at various time points (0, 0.5, 1, 1.5, 2.5h) following independent labeling of both the Beta-TC-6 and MC-38 cell lines with the ^{18}F-FDG radioactive isotope in increasing cell amounts (3.125×10^5, 6.25×10^5, 1.25×10^6, and 2.5×10^6 cells; n = 4) (Figures 5A and 5B). Following repeated measures of the initial radioactivity in the different cell numbers of both cell lines, standardized classification bins with varying ranges for distinct cell numbers were established (Table S1). This classification bin was eventually programmed into the DL algorithm to allow for identification of cell number based on the predicted radioactivity from the algorithm. The MC-38 cell line displayed greater average radioactive isotope uptake across all cell numbers in comparison to the Beta-TC-6 cell line, due to the greater relative cellular metabolic activity (Figure 5A). There is a distinct difference between the initial radioactivity values of the various cell numbers across both cell lines, indicating a lack of overlap in these values. This renders the unique utility of initial radioactivity values of a sample of labeled cells and highlights the importance of resolving the predicted radioactivity from the algorithm to an initial time point to predict cell number using PET/MRI. The common pattern of decay within each number of cells indicates that if the initial time point of cell labeling is known, as well as the current time of the scan, then the half-life decay equation for the radioactive isotope under study, in this case ^{18}F, can be used to calculate the original radioactivity of the sample in the PET/MRI scan regardless of the time of scan. This can be done by calculating the minutes that have passed since the initial time point of labeling and using the predicted radioactivity from the DL algorithm at the time of scan and arranging for the half-life equation to calculate the initial radioactivity using the values of time (minutes passed) and final radioactivity. Importantly, the calculation of the original radioactivity to classify the cell number in the PET/MRI scan is only useful if an accurate prediction of the current radioactivity in the scan is achieved by a well-trained and calibrated DL algorithm. Once the radioactivity at the original time point of labeling is calculated for a sample of cells, a classification of cell number is made by sorting the predicted, time-corrected radioactivity into one of the classification bins based on the range of radioactivity in which the value falls.

Training of 3D CNN before and after image rescale & reconstruction

Upon initial training of the algorithm without reconstruction of the raw PET image data, MSSE values for all ranges of radioactivity value predictions were high, with minimal decrease in MSSE values across increasing

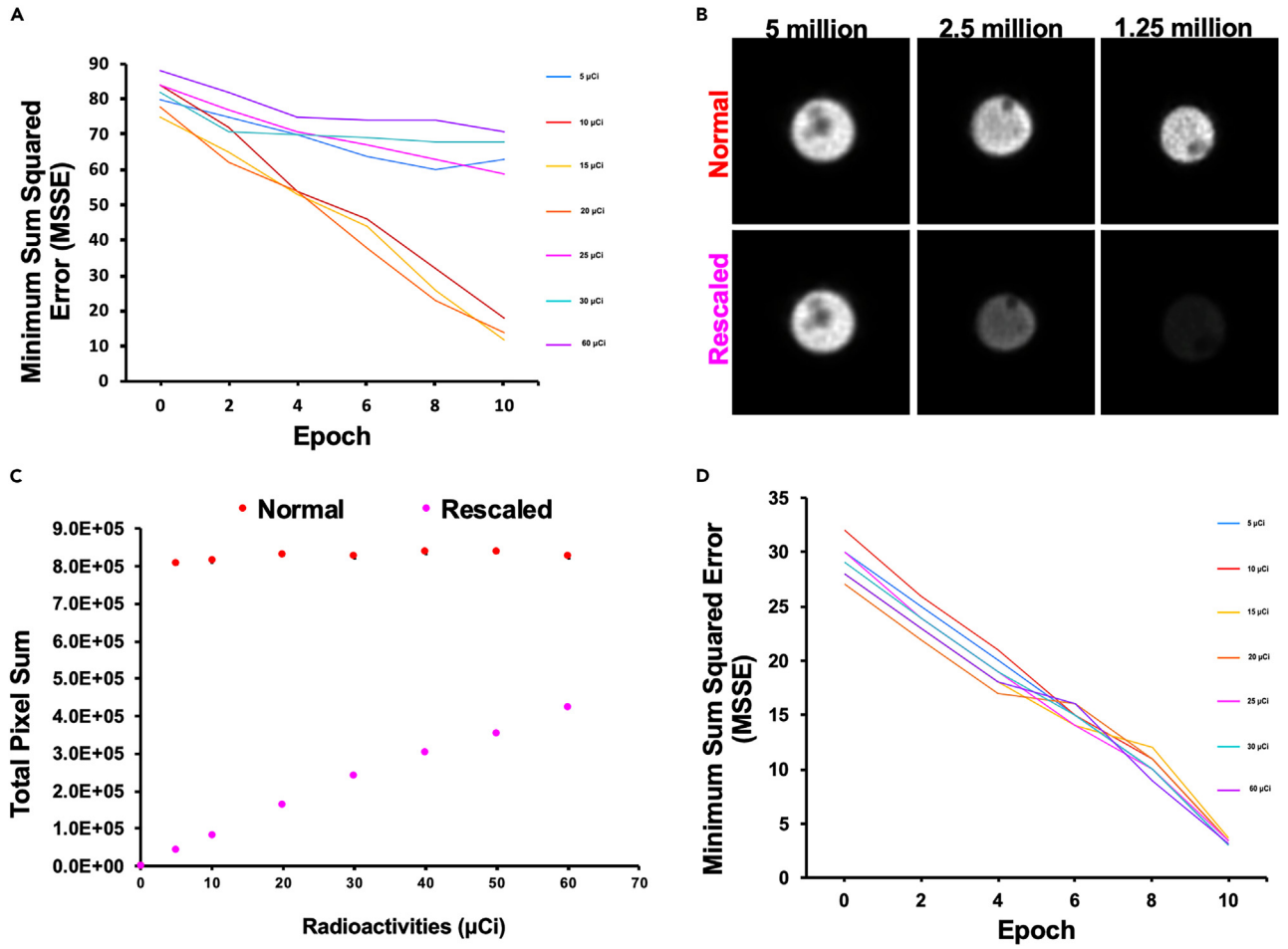


Figure 6. Application of rescale equation in the preprocessing of images and the impact on training of 3D CNN

(A) Minimum sum squared error (MSSE) among ten epochs for 5, 10, 15, 20, 25, 30, 60 µCi radioactivity labeled cell sample images before rescale equation was applied to algorithm.

(B) (Top Row) Normal 3D PET images of 5, 2.5, and 1.25 million cells before rescale equation was applied in preprocessing (Bottom Row) Rescaled 3D PET images after rescale equation applied in preprocessing.

(C) Total pixel sum of images representing a range of 0–70 µCi before (normal) and after (rescaled) rescale equation was applied to 3D PET images used for training the algorithm.

(D) MSSE after rescale equation was applied to 3D PET images of labeled cells of 5, 10, 15, 20, 25, 30, 60 µCi radioactivity.

epochs of training (Figure 6A). More notably, the MSSE for radioactivity predictions on the ends of the spectrum of the range of radioactivities used in the study, such as that for less than 10µCi or greater than 20 µCi, remained flat without any significant decrease as the number of training epochs increased. Although for the middle range of radioactivity, between 10 and 20 µCi, the MSSE decreased among the iterations of training, the greater overall values of MSSE indicated that the neural network was not performing with a great degree of accuracy as compared to when the image data were reconstructed. This was due to the intrinsic properties of CNN, which relies more heavily on image texture to establish a proper gradient for learning and decision making in the prediction process. As a result, the images were evaluated with raw format to detect a visual difference between increasing cell numbers and this increasing radioactivity, upon which no detectable qualitative difference between the images was found (Figure 6B, top row). Further evaluation of the DICOM image files and their respective headers revealed that the images were being intrinsically altered by the scanner through a scaling factor which standardizes the maximum and minimum pixel intensity values across all images, and this was masking the true nature and texture of the images. Computational evaluation of the maximum and minimum intensities of images of different cell numbers and radioactivities confirmed this hypothesis, with the maximum intensity remaining around 65535 for all images despite their varying radioactivities (Figure 6C). This prompted the establishment of a rescale equation (Equation 1) to undo the scaling done

by the scanner through the DICOM header for each image volume, to unmask the true nature and pixel values for the dataset. Upon rescaling the images to their true maximum intensities reflective of their respective radioactivity values/labeled cell numbers, a pattern of increasing maximum intensity and total pixel sum for increasing radioactivity and labeled cell numbers was revealed (Figure 6C). This pattern was also apparent visually after rescaling, with increasing pixel intensity as cell number increased (Figure 6B). Upon training of the CNN after the rescale equation was applied to the images in preprocessing, similar decrease in MSSE upon increasing epochs of training was found, indicating increasing accuracy of algorithm prediction of radioactivity because of revealing the true linear pattern of increasing pixel intensity, which is indicative of increased radioactivity, with increased labeled cells in the images (Figure 6D).

Graft cell number prediction

Following initial training of the 3D CNN algorithm on the *in vitro* phantom models for adequate predictive quantification of all potential ranges of radioactivity in a 3D PET scan of ^{18}F-FDG-labeled cells, the beta-TC6 cells were transplanted in varying cell counts under the left kidney capsule of mice (n = 9) (Figure 7A). PET/MRI verified the cells were transplanted in the intended location and maintained a detectable level of radioactivity after transplantation. These *in vivo* images were plugged into the CNN algorithm for testing of the model's prediction on *in vivo* transplant images as it was trained on phantoms representing the labeled cells. It is important to note that cells could be lost during the transplantation process, as vestiges may remain in the syringe or unforeseen errors can cause a reduced number of cells than intended to be transplanted. Hence, it is critical for CNN to predict the radioactivity of the transplant cells with a high degree of accuracy and classify the transplanted cell number based on this quantification.

For *in vivo* monitoring purposes, the acquired *in vivo* data from different time points were plugged into the trained 3D CNN radioactivity prediction model within the mice post-transplantation and subsequent estimation of transplanted cell number in the region based on the predicted radioactivity determined by the algorithm. For *in vitro* phantom models of both cell lines, the algorithm performed with a strong degree of accuracy in comparison with ground truth values. For the *in vivo* model predictions, the accuracy of the estimation of transplanted cell number decreased as per the lower ICC values seen for the maximum and minimum cell numbers (Figure 7B). This discrepancy in predictive accuracy observed between the 3D CNN's predictions on *in vitro* as compared to the *in vivo* data results from two major confounding factors: there are more steps in preparation and thus chance for cell loss. The preparation process of the *in vitro* phantoms is straightforward and does not require a great deal of transfer of cells from flask to tubes and finally to surgical tubing as is required in the *in vivo* cell transplantation process. The direct transfer of the cells in their estimated amounts from the flask to the phantoms in which they will be imaged reduces the chances that intended cells may be mistakenly lost or uncontrollably left in one of the many constructs in which they are transferred and processed as is done in the *in vivo* transplantation process. Furthermore, for the *in vivo* model, the reduced ICC values observed in cell number prediction result from a difference between the number of cells that the experimenters believe they have transplanted in the mouse and the number of cells that were injected through the surgical tubing under the left kidney capsule and remained in the mouse for the duration of imaging and thereafter (Figure 7B). Hence, this discrepancy for *in vivo* models is crucial for support of the initial hypothesis and goal of the study which aims to highlight the fact that the experimenters require assistance in quantifying the actual number of transplanted cells due to the possible chances for cell or graft loss during the preparation and operation stages of the transplantation process. Here, the algorithm which has been trained, tested, and validated in its capacity to predict the radioactivity within 3D PET/MRI scans of similar format and similarly labeled cells can more accurately define the number of cells that are transplanted in the animal models based on its ability to directly quantify the radioactivity in the image based on previous training and associations on similarities in image texture and their associated radioactivity values. This cannot be done automatically by the human experimenter alone when evaluating a 3D PET/MRI scan of transplanted labeled cells. Hence, the inconsistency observed between the experimenter's estimation of how many cells they have transplanted compared to the 3D CNN's predictive analysis on the actual radioactivity and number of cells remaining under the kidney capsule at the time of scanning highlights the importance of training a DL algorithm to help accurately analyze such scans for evaluation of transplanted cell number as indicated by radioactivity in the ROI.

DISCUSSION

There are numerous clinical PET/MRI applications at both the pre-clinical and clinical stage. However, numerous studies have compared PET/MRI with PET/CT and have emphasized that experience with

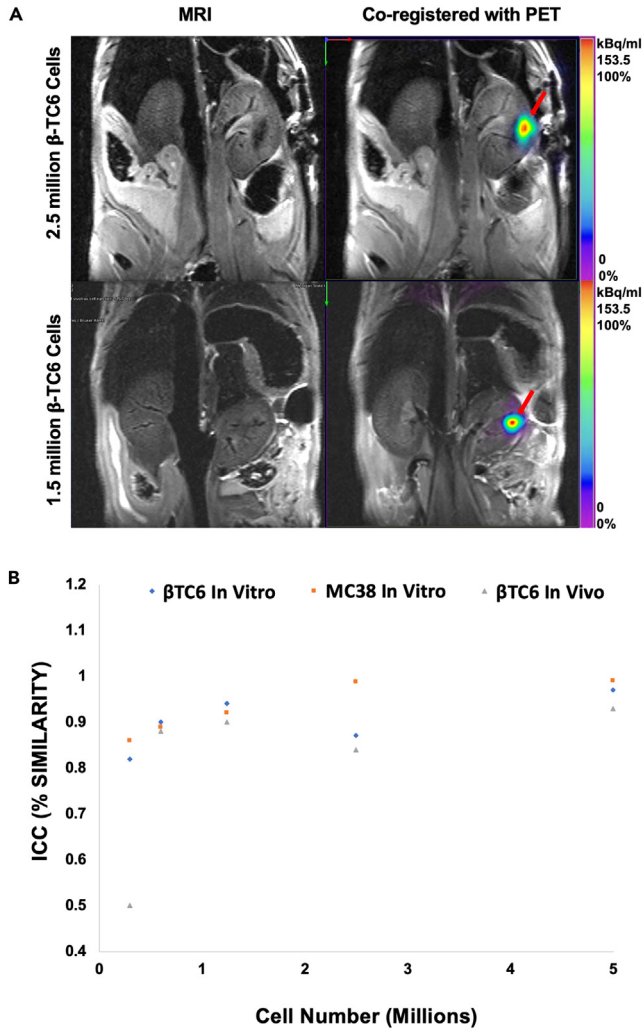


Figure 7. *In vivo* simultaneous PET/MRI imaging and cell number prediction of cell transplantation under left kidney capsule of mice

(A) (Left) T2WI MRI image of transplanted labeled beta-TC6 cells under left kidney capsule of mice (red arrow) (Right) MRI image coregistered to PET scan of transplanted labeled beta-TC6 cells shows PET signal under the left kidney capsule of the mice (red arrow) (Top = 2.5 million cells, Bottom = 1.5 million cells). Scale bar of PET images is 0-153.5 kBq/ml, same for both groups.

(B) Intraclass correlation coefficient (ICC) indicating percent similarity of the algorithms classification of cell number based on input images to the true cell number as calculated manually for *in vitro* phantom images of both BTC6 cell line images (purple) and MC38 cell line images (blue) and *in vivo* images of transplanted BTC6 cell line images (pink).

PET/MRI is limited, and further research should be performed on this modality to fully understand its capabilities within the realm of medical imaging and intervention.[8,9,13,17] Currently, PET/MRI has been used for imaging of tumors in patients with oncologic diagnoses which include anal cancer, breast cancer, colon cancer, esophageal cancer, leukemia, lymphoma, and pancreatic cancer.[8] In addition, it has also been used for patients with disorders in the cardiovascular and central nervous system, as well as pediatric patients in general.[1,10,26]

The results of the study successfully highlight the effectiveness and accuracy of a dynamically trained 3D CNN in predicting transplanted cell number in simultaneous PET/MRI scans. It also portrays the capability of a combined algorithm, using *k-means++* clustering, to automatically segment transplanted cell ROI from the simultaneous PET/MRI image. This was achieved through combining *K-means++*, an unsupervised ML algorithm for segmentation of the scans,[24,25] with a 3D CNN, which was used for automated

PET signal quantification and subsequent cell number quantification. Through ICC-based statistical analysis of algorithm performance in comparison with manual quantification and estimation, we have shown that the algorithm quantifies the ROI with a great degree of accuracy compared to the standard dose calibrator device used in the lab and with superior accuracy compared to human counterparts who are unable to predict the radioactivity of the PET signal within the image. A similar, high degree of accuracy was seen with quantification of the cell number using the predicted radioactive dose from the 3D CNN, as indicated by the ICC scores greater than 90% for nearly all possible cell numbers in both *in vitro* and *in vivo* images. Regarding segmentation, the algorithm was capable of segmenting islands of pixels from the entire pixel field within a PET image, a task which is prone to error and is often overlooked by human analysts.

The engineering and training of a highly accurate and reliable DL model in this study was contingent upon the generation of representative, dynamic, and uniform data.[27] Often, in DL training, gathering such data is difficult, and their setup and retrieval are often prone to errors. The number of acquired data points is important for successful training of DL algorithms, with an exponential correlation between increasing number of available training data and the consequent decrease in loss value (and thus increase in performance) of the model. Here, the challenge was generating enough training data that would represent the spatiotemporal distribution of signal when labeled cells were transplanted in mice. Of course, transplanting and sacrificing over 100 mice were both unreasonable and unethical, so a new method of generating a large amount of reliable and representative training data was devised. After confirming that the metabolism of ^{18}F-FDG was independent of the placement of the cells *in vitro* or *in vivo*, and that the spatiotemporal nature of the signal was unchanged between *in vitro* and *in vivo* models of labeled cells, hundreds of phantoms were able to be generated with the same image dimensions as the *in vivo* dataset for which the algorithm would be used to monitor cell transplants. This large volume of data represented the dynamic nature of signal ROIs in simultaneous PET/MRI which corresponded to different radioactivity (and thus different cell numbers) and thus allowed for robust training of the 3D CNN automatic for signal quantification. The training of the model used for radioactivity prediction was based on phantoms generated with increasing amounts of cells labeled with ^{18}F-FDG, which permitted the collection of data with increasing image texture-related parameters such as total pixel sum and integrated density among images with increasing radioactive dose and cell numbers. After observing a linear pattern between increasing ^{18}F-FDG in the form of labeled cells and the corresponding image texture in PET, we trained the CNN on more than 150 raw PET image volumes with their associated radioactivity values as measured by the dose calibrator device. The network was trained on recognizing the corresponding initial radioactivity to each PET image volume which consisted of 3D spatial distribution of pixels of a particular intensity and pattern. The initial radioactivity which was recorded at the time of scanning was recorded using the dose calibrator device and used in input/output training by treating the 3D pixel network of a single PET image volume as the input and its corresponding, measured radioactivity value as the output value. Upon training of the CNN to predict the initial radioactivity from PET scans, statistical evaluation of the algorithm performance in comparison with the dose calibrator device yielded excellent (0.9+) ICC scores which indicate a high degree of accuracy for prediction of the radioactivity from the images. The positive ICC score of comparison of the DL algorithm with the physical dose calibrator device indicated that the CNN could quantify the radioactivity within an image volume with high spatiotemporal accuracy.

To this effect, the algorithm was capable of "correcting" (back-calculating) the radioactivity to an initial time point when cells were labeled using the rearranged time-decay equation for ^{18}F (Equation 2); this corrected radioactivity value indicated the number of cells. This was due to the observation of a linear trend in increasing radioactivity with increasing cell number during the cell characterization phase of our study, in which different cell lines (BTC6 and MC38) metabolized different amounts of ^{18}F-FDG initially yet retained the same linear trend of increasing radioactivity with increasing cell number.[28,29] This permitted CNN to predict the cell number from the time-corrected radioactivity value in a cell line-dependent manner. We have chosen the cell classification bin due to the overlap in radioactivity values between different numbers of cells from distinct cell lines. This results from different labeling efficiency between cell lines due to varying cell metabolism (hence, varying propensity toward glucose uptake). This is further complicated by the inability of a CNN to distinguish the exact cell line solely based on the PET data generated in this study, and therefore it would not be able to predict cell number accurately among different cell lines. Classification bins were programmed into the inference script of the neural network for estimation of cell number from the predicted radioactivity, with two different sets of classification bins for the two different cell lines we evaluated in the study. It can be inferred that in cell-number prediction of a future, unseen cell line with this algorithm, initial characterization of the cell line through measurement of ^{18}F-FDG uptake radioactivity

per different cell numbers within that cell line is necessary to create the cell line-specific classification bins that the CNN will use to make its estimation. This is similar to initial calibration that is performed for various DL and ML algorithms in the realm of biomedical imaging.

In the training of the algorithm on a dynamic array of possible radioactivities it may encounter in a future setting allowed for the augmentation of a dose calibrator, a device used in clinic to measure radioactivity in samples prior to or after injection, preceding its end task in using the quantified, predicted radioactivity that was detected in the image to classify the cell number in the region based on the developed classification bins. Remarkably, the trained model was able to augment the dose calibrator by accurately predicting (quantifying) the radioactive dose in the ROI for all new images that were fed into the neural network, regardless of their *in vivo* or *in vitro* nature. This foundational, independent role of the 3D CNN in accurately quantifying the radioactivity dose in an encountered sample at the time of imaging provides a useful tool for the field of PET/MRI alone. Although here it is primarily used with the endpoint goal of cell number quantification using the predicted radioactivity, it can also be used to quantify the uptake of radiolabeled drugs and measure response to treatment in future applications of this algorithm. These applications pertain heavily to the forthcoming and rapidly evolving field of precision medicine, particularly in respect to the development of novel radiolabeled nanodrugs and tracers for theranostic imaging. In this regard, the algorithm provides a novel paradigm for AI algorithms to automatically segment and analyze an ROI from PET or simultaneous PET/MRI imaging for a wide variety of domains within the field. As mentioned earlier, this permits the establishment of methods of quantification, assessment, and analysis that are not prone to human error and bias and can act as a standardized means of monitoring novel treatments and therapies with PET and simultaneous PET/MRI.[30]

Limitations of the study

There are several limitations of our study. Our method faces one limitation as it did not label the cells with dual-modality PET/MRI probes.[31,32] The integration of PET and MRI modalities into a single hybrid imaging system has been demonstrated to synergistically compensate for the limitations of each modality.[33] In our current study, MRI was used for anatomical reference, not for segmentations, although this does not compromise the quality of segmentation performed solely based on PET data, which adequately reflected the location of the cells within the constructs, nor does it have any impact on quantification of radioactivity and cell number. However, in future studies we may devise methods to label cells with all-in-one dual-modality probes that permit the incorporation of the MRI data into the pipeline for potentially better segmentation, prediction, and quantification. Another limitation of our study is that the current combined algorithm was only tested for ^{18}F-FDG; we will test other radiotracers in our future studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cells
  - Animals
- METHOD DETAILS
  - Cell labeling with ^{18}F-FDG
  - Phantom preparation and radioactivity quantification
  - *In vivo* cell transplantationin mice
  - PET/MRI
  - PET image preprocessing and rescale
  - 3D CNN architecture, training and cell number classification
  - 2D CNN architecture
  - Gradient-boosted decision tree architecture
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107083.

ACKNOWLEDGMENTS

MSU has a research agreement with Bruker BioSpin. The work of A.S., S.N., and P.W. was supported in part by NIH grants R03EB028349-01A1 and R21AI159928-01. The work of R.W. and G.W.W. was supported in part by NIH grants R01GM126189 and R01AI164266.

AUTHOR CONTRIBUTIONS

Conceptualization, J.F. and P.W.; Methodology, H.H., R.W., J.F. G.W.W., and P.W.; Software, H.H., R.W., and P.W.; Investigation, H.H., R.W., A.S., C.L.M., S.N., N.R., D.B., E.G., N.T., J.F., and P. W.; Data Curation, H.H. and R.W.; Writing – Original Draft, H.H., R.W., and P.W.; Writing – Review & Editing, R.W., C.L.M., N.R., G.W.W., J.F., and P.W.; Resources, A.A., A.M., and K.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: September 24, 2022
Revised: February 10, 2023
Accepted: June 6, 2023
Published: June 9, 2023

REFERENCES

1. Schlosser, T., Nensa, F., Mahabadi, A.A., and Poeppel, T.D. (2013). Hybrid MRI/PET of the heart: a new complementary imaging technique for simultaneous acquisition of MRI and PET data. Heart *99*, 351–352. https://doi.org/10.1136/heartjnl-2012-302740.

2. Mueller, W.P., Melzer, H.I., Schmid, I., Coppenrath, E., Bartenstein, P., and Pfluger, T. (2013). The diagnostic value of 18F-FDG PET and MRI in paediatric histiocytosis. Eur. J. Nucl. Med. Mol. Imaging *40*, 356–363. https://doi.org/10.1007/s00259-012-2278-6.

3. Jadvar, H., and Colletti, P.M. (2014). Competitive advantage of PET/MRI. Eur. J. Radiol. *83*, 84–94. https://doi.org/10.1016/j.ejrad.2013.05.028.

4. Tatsumi, M., Yamamoto, S., Imaizumi, M., Watabe, T., Kanai, Y., Aoki, M., Kato, H., Shimosegawa, E., and Hatazawa, J. (2012). Simultaneous PET/MR body imaging in rats: initial experiences with an integrated PET/MRI scanner. Ann. Nucl. Med. *26*, 444–449. https://doi.org/10.1007/s12149-012-0581-8.

5. Jung, J.H., Choi, Y., and Im, K.C. (2016). PET/MRI: technical challenges and recent advances. Nucl. Med. Mol. Imaging *50*, 3–12. https://doi.org/10.1007/s13139-016-0393-1.

6. Musafargani, S., Ghosh, K.K., Mishra, S., Mahalakshmi, P., Padmanabhan, P., and Gulyás, B. (2018). PET/MRI: a Frontier in era of complementary hybrid imaging. Eur. J. Hybrid Imaging *2*, 12. https://doi.org/10.1186/s41824-018-0030-6.

7. Brix, G., Lechel, U., Glatting, G., Ziegler, S.I., Münzing, W., Müller, S.P., and Beyer, T. (2005). Radiation exposure of patients undergoing whole-body dual-modality 18F-FDG PET/CT examinations. J. Nucl. Med. *46*, 608–613.

8. Drzezga, A., Souvatzoglou, M., Eiber, M., Beer, A.J., Fürst, S., Martinez-Möller, A., Nekolla, S.G., Ziegler, S., Ganter, C., Rummeny, E.J., and Schwaiger, M. (2012). First clinical experience with integrated whole-body PET/MR: comparison to PET/CT in patients with oncologic diagnoses. J. Nucl. Med. *53*, 845–855. https://doi.org/10.2967/jnumed.111.098608.

9. Schwenzer, N.F., Schraml, C., Müller, M., Brendle, C., Sauter, A., Spengler, W., Pfannenberg, A.C., Claussen, C.D., and Schmidt, H. (2012). Pulmonary lesion assessment: comparison of whole-body hybrid MR/PET and PET/CT imaging–pilot study. Radiology *264*, 551–558. https://doi.org/10.1148/radiol.12111942.

10. Boss, A., Bisdas, S., Kolb, A., Hofmann, M., Ernemann, U., Claussen, C.D., Pfannenberg, C., Pichler, B.J., Reimold, M., and Stegger, L. (2010). Hybrid PET/MRI of intracranial masses: initial experiences and comparison to PET/CT. J. Nucl. Med. *51*, 1198–1205. https://doi.org/10.2967/jnumed.110.074773.

11. Tarantal, A.F., Lee, C.C.I., Kukis, D.L., and Cherry, S.R. (2013). Radiolabeling human peripheral blood stem cells for positron emission tomography (PET) imaging in young rhesus monkeys. PLoS One *8*, e77148. https://doi.org/10.1371/journal.pone.0077148.

12. Huang, J., Lee, C.C., Sutcliffe, J.L., Cherry, S.R., and Tarantal, A.F. (2008). Radiolabeling rhesus monkey CD34+ hematopoietic and mesenchymal stem cells with 64Cu-pyruvaldehyde-bis(N4-methylthiosemicarbazone) for microPET imaging. Mol. Imaging *7*, 1–11.

13. Beiderwellen, K.J., Poeppel, T.D., Hartung-Knemeyer, V., Buchbender, C., Kuehl, H., Bockisch, A., and Lauenstein, T.C. (2013). Simultaneous 68Ga-DOTATOC PET/MRI in patients with gastroenteropancreatic neuroendocrine tumors: initial results. Invest. Radiol. *48*, 273–279. https://doi.org/10.1097/RLI.0b013e3182871a7f.

14. Pagel, J.M., Matthews, D.C., Appelbaum, F.R., Bernstein, I.D., and Press, O.W. (2002). The use of radioimmunoconjugates in stem cell transplantation. Bone Marrow Transplant. *29*, 807–816. https://doi.org/10.1038/sj.bmt.1703524.

15. Gopal, A.K., and Winter, J.N. (2009). Radioimmunoconjugates in hematopoietic stem cell transplantation. Cancer Treat Res. *144*, 299–315. https://doi.org/10.1007/978-0-387-78580-6_13.

16. Nose, N., Nogami, S., Koshino, K., Chen, X., Werner, R.A., Kashima, S., Rowe, S.P., Lapa, C., Fukuchi, K., and Higuchi, T. (2021). [18F] FDG-labelled stem cell PET imaging in different route of administrations and

multiple animal species. Sci. Rep. *11*, 10896. https://doi.org/10.1038/s41598-021-90383-4.

17. Kohan, A.A., Kolthammer, J.A., Vercher-Conejero, J.L., Rubbert, C., Partovi, S., Jones, R., Herrmann, K.A., and Faulhaber, P. (2013). N staging of lung cancer patients with PET/MRI using a three-segment model attenuation correction algorithm: initial experience. Eur. Radiol. *23*, 3161–3169. https://doi.org/10.1007/s00330-013-2914-y.

18. Wei, X., Zhou, L., Chen, Z., Zhang, L., and Zhou, Y. (2018). Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. BMC Med. Inform. Decis. Mak. *18*, 111. https://doi.org/10.1186/s12911-018-0693-8.

19. Ullah, W., Siddique, I., Zulqarnain, R.M., Alam, M.M., Ahmad, I., and Raza, U.A. (2021). Classification of arrhythmia in heartbeat detection using deep learning. Comput. Intell. Neurosci. *2021*, 2195922. https://doi.org/10.1155/2021/2195922.

20. Franks, J.M., Martyanov, V., Wang, Y., Wood, T.A., Pinckney, A., Crofford, L.J., Keyes-Elstein, L., Furst, D.E., Goldmuntz, E., Mayes, M.D., et al. (2020). Machine learning predicts stem cell transplant response in severe scleroderma. Ann. Rheum. Dis. *79*, 1608–1615. https://doi.org/10.1136/annrheumdis-2020-217033.

21. Shouval, R., Labopin, M., Bondi, O., Mishan-Shamay, H., Shimoni, A., Ciceri, F., Esteve, J., Giebel, S., Gorin, N.C., Schmid, C., et al. (2015). Prediction of allogeneic hematopoietic stem-cell transplantation mortality 100 Days after transplantation using a machine learning algorithm: a European group for blood and marrow transplantation acute leukemia working party retrospective data mining study. J. Clin. Oncol. *33*, 3144–3151. https://doi.org/10.1200/JCO.2014.59.1339.

22. McCall, M., and Shapiro, A.M.J. (2012). Update on islet transplantation. Cold Spring Harb. Perspect. Med. *2*, a007823. https://doi.org/10.1101/cshperspect.a007823.

23. Wang, P., Yigit, M.V., Ran, C., Ross, A., Wei, L., Dai, G., Medarova, Z., and Moore, A. (2012). A theranostic small interfering RNA nanoprobe protects pancreatic islet grafts from adoptively transferred immune rejection. Diabetes *61*, 3247–3254. https://doi.org/10.2337/db12-0441.

24. Hayat, H., Sun, A., Hayat, H., Liu, S., Talebloo, N., Pinger, C., Bishop, J.O., Gudi, M., Dwan, B.F., Ma, X., et al. (2021). Artificial intelligence analysis of magnetic particle imaging for islet transplantation in a mouse model. Mol. Imaging Biol. *23*, 18–29. https://doi.org/10.1007/s11307-020-01533-5.

25. Sun, A., Hayat, H., Liu, S., Tull, E., Bishop, J.O., Dwan, B.F., Gudi, M., Talebloo, N., Dizon, J.R., Li, W., et al. (2021). 3D in vivo magnetic particle imaging of human stem cell-derived islet organoid transplantation using a machine learning algorithm. Front. Cell Dev. Biol. *9*, 704483. https://doi.org/10.3389/fcell.2021.704483.

26. White, J.A., Rajchl, M., Butler, J., Thompson, R.T., Prato, F.S., and Wisenberg, G. (2013). Active cardiac sarcoidosis: first clinical experience of simultaneous positron emission tomography–magnetic resonance imaging for the diagnosis of cardiac disease. Circulation *127*, e639–e641. https://doi.org/10.1161/CIRCULATIONAHA.112.001217.

27. Ihme, M., Chung, W.T., and Mishra, A.A. (2022). Combustion machine learning: principles, progress and prospects. Prog. Energy Combust. Sci. *91*, 101010. https://doi.org/10.1016/j.pecs.2022.101010.

28. Chan, L.W., Hapdey, S., English, S., Seidel, J., Carson, J., Sowers, A.L., Krishna, M.C., Green, M.V., Mitchell, J.B., and Bacharach, S.L. (2006). The influence of tumor oxygenation on (18)F-FDG (fluorine-18 deoxyglucose) uptake: a mouse study using positron emission tomography (PET). Radiat. Oncol. *1*, 3. https://doi.org/10.1186/1748-717X-1-3.

29. Buck, A.K., and Reske, S.N. (2004). Cellular origin and molecular mechanisms of 18F-FDG uptake: is there a contribution of the endothelium? J. Nucl. Med. *45*, 461–463.

30. Hope, T.A., Fayad, Z.A., Fowler, K.J., Holley, D., Iagaru, A., McMillan, A.B., Veit-Haiback, P., Witte, R.J., Zaharchuk, G., and Catana, C. (2019). Summary of the first ISMRM-SNMMI workshop on PET/MRI: applications and limitations. J. Nucl. Med. *60*, 1340–1346. https://doi.org/10.2967/jnumed.119.227231.

31. Lahooti, A., Sarkar, S., Laurent, S., and Shanehsazzadeh, S. (2016). Dual nano-sized contrast agents in PET/MRI: a systematic review. Contrast Media Mol. Imaging *11*, 428–447. https://doi.org/10.1002/cmmi.1719.

32. Wei, X., Zhao, H., Huang, G., Liu, J., He, W., and Huang, Q. (2022). ES-MION-Based dual-modality PET/MRI probes for acidic tumor microenvironment imaging. ACS Omega *7*, 3442–3451. https://doi.org/10.1021/acsomega.1c05815.

33. Garcia, J., Tang, T., and Louie, A.Y. (2015). Nanoparticle-based multimodal PET/MRI probes. Nanomedicine *10*, 1343–1359. https://doi.org/10.2217/nnm.14.224.

34. Zhao, H., Hayat, H., Ma, X., Fan, D., Wang, P., and Moore, A. (2020). Molecular imaging and deep learning analysis of uMUC1 expression in response to chemotherapy in an orthotopic model of ovarian cancer. Sci. Rep. *10*, 14942. https://doi.org/10.1038/s41598-020-71890-2.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
^{18}F-FDG	Cardinal Health	https://www.cardinalhealth.com/
Agarose	Sigma-Aldrich	A9539
Deposited data		
Raw and analyzed data	This paper; Mendeley Data	https://doi.org/10.17632/995vjkhm68.1
Codes for K-mean ++, 2D CNN, 3D CNN and GBDT	This paper; Mendeley Data	https://doi.org/10.17632/995vjkhm68.1
Experimental models: Cell lines		
Beta-TC-6 cells	ATCC	RRID: CVCL_0605
MC38 cells	ATCC	RRID: CVCL_B288
Experimental models: Organisms/strains		
Mouse: NOD/scid	The Jackson Laboratory	RRID: IMSR_JAX:001303
Software and algorithms		
Python version 3.7	Python Software Foundation	https://www.python.org
ParaVision 360 1.1 and 3.1	Bruker, Billerica MA	https://www.bruker.com/en/products-and-solutions/preclinical-imaging/paravision-360.html
SPSS statistical software 28.0.1.1	IBM, Armonk, NY	https://www.ibm.com/products/spss-statistics
Other		
BioSpec 70/30 MRI scanner with a PET insert	Brucker, Billerica, MA	https://www.bruker.com/en/products-and-solutions/preclinical-imaging/mri.html
Dose calibrator	Capintec, Florham Park, NJ	CRC-55tR
J750 3D printer	Stratasys Ltd., Eden Prairie, MN	https://www.stratasys.com/en/resources/blog/j750-3d-printer-introduction/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ping Wang (wangpin4@msu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All training and testing imaging data have been deposited at Mendeley. The DOIs are listed in the key resources table.

All original codes for K-mean ++, 2D CNN, 3D CNN and GBDT have been deposited at Mendeley. The DOIs are listed in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cells

Beta-TC-6 cells were cultured in 2.8 mM glucose DMEM (Gibco, Thermo Fisher Scientific, MA) with 10% FBS and 1% penicillin/streptomycin in an incubator at 37°C, 5% CO_2. Cells were cultured to a confluence of approximately 10 million cells per 500 mL flask. MC38 cells were cultured in DMEM (Gibco, Thermo Fisher Scientific, MA) with 10% FBS (Thermo Fisher Scientific, MA) and 1% penicillin/streptomycin (Gibco, Thermo Fisher Scientific, MA) in an incubator at 37°C, 5% CO_2.

Animals

All animal experiments were conducted according to the institutional guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee (IACUC) at Michigan State University. All NOD/scid immunodeficient mice were obtained from the Jackson Laboratories (Bar Harbor, ME) and housed at animal housing facilities with a 12-h light-dark cycle at Michigan State University. Campus Animal Resources at Michigan State University provides veterinary care, daily husbandry, and health checks. Female NOD/scid mice (n = 6, 10 weeks old) were used for *in vivo* experiments.

METHOD DETAILS

Cell labeling with ^{18}F-FDG

Both cell lines were labeled with ^{18}F-FDG at a dose of 6 mCi per flask and incubated for 1 h. Following incubation, cells were washed three times with PBS to remove free ^{18}F-FDG and prepared for digestion with trypsin for phantom preparation and transplantation.

Phantom preparation and radioactivity quantification

Cells were digested with 8mL of 0.25% Trypsin for 2 min and quenched with 12 mL 2.8 mM glucose DMEM (Gibco, Thermo Fisher Scientific, MA) with 10% FBS and 1% penicillin/streptomycin for a final volume of 20 mL/10 M (M = million) cells suspended in culture medium per flask. Aliquots in increments of 0.625, 1.25, 2.5, 5, and 10 mL corresponding to approximate cell number amounts of 3.125×10^5, 6.25×10^5, 1.25×10^6, 2.5×10^6, and 5×10^6 cells, respectively, were transferred to centrifuge tubes. The tubes were centrifuged at a rate of 1000 RPM for 5 min at a temperature of 20°C. During centrifugation, prepared 1% agarose gel was warmed for approximately 30 s in the microwave (until liquefied). The gel was held in a 50 mL tube placed in a beaker of warm water to prevent solidification of the gel for duration of the phantom preparation process. Following centrifugation, supernatant was removed carefully from each tube using sterile pipette technique. Once all supernatant was removed from the tubes and only cell pellets remained, the cells were resuspended and mixed in 200μL of the 1% agarose gel using pipette pumping technique and transferred to the 3D-printed phantoms (Figure S2) after mixing by trituration until homogeneous. Phantoms were labeled with their respective cell amounts and subsequently placed one at a time into the dose calibrator device for immediate quantification of radioactivity. Radioactivity values were recorded in μCi for each different cell number phantom and the values were corrected to the initial timepoint of labeling for generation of a standard curve of cell number corresponding to different radioactivity for the cell line being studied.

In vivo cell transplantationin mice

Labeled cells were detached and collected with 0.25% trypsin. Female NOD/scid immunodeficient mice (n = 6, 10 weeks old, Jackson Laboratory, Bar Harbor, ME) were anesthetized with 2% isoflurane. An incision was made to expose the left kidney of the mouse, then a catheter needle was inserted underneath the kidney capsule and cells labeled with ^{18}F-FDG were transplanted.

PET/MRI

Images were acquired on a Biospec 70/30 with a PET insert using Paravision 1.1 and 3.1 (Bruker, Billerica, MA). MRI images of phantoms were acquired using an 86-mm transmit/receive volume coil: T_2 map parameters were: multi-spin multi-echo sequence, TR/TE: 2200/8.4, field of view 60 × 30 mm, resolution 200 × 200 × 800 μm, 15 slices, 12 echo images acquired (fewer points were used for generating maps, depending on signal becoming equivalent to the noise), acquisition time 5min30s. PET images were acquired for 10 min, and were reconstructed using a calibrated MLEM (maximum likelihood expectation maximization)

algorithm with a resolution of 0.5 mm and 16 iterations, and corrections for scatter, randoms, decay and partial volume. Phantom images were not attenuation corrected.

Mice were positioned on top of a 4 cm 4-channel array receive coil and an 86-mm volume coil was used to transmit the RF signal. T_2* weighted image: 2D T_1 FLASH, TR/TE: 200/8 ms, field of view 30 × 30mm, resolution 200 × 200 × 500 μm, 6 slices, flip angle 30°, 16 averages, acquisition time 4min3s. T_2 weighted image: 2D-TurboRARE, TR/TE: 2200/40 ms, RARE factor 8, field of view 30 × 30 mm, resolution 200 × 200 × 500 μm, 6 slices, 4 averages, acquisition time 2min38s. T2map: multi-spin multi-echo sequence, TR/TE: 2200/8.4, field of view 30 × 30mm, resolution 200 × 200 × 500 μm, 12 echo images acquired (fewer points were used for generating maps, depending on signal becoming equivalent to the noise), acquisition time 5min30s. A 3D FISP image with 200 × 200 × 400 μm resolution was used for attenuation correction of the PET signal for the mouse images.

Mouse PET images were acquired for 30 min, and were reconstructed using a calibrated MLEM method with a resolution of 0.5 mm and 48 iterations, and corrections for scatter, randoms, decay, and partial volume.

PET image preprocessing and rescale

Raw, unregistered 3D PET DICOM image sequences of voxel dimensions (sized $_{0.5mm}$) [z, x, y] mm = [300 × 160 × 160] were resampled to reduced dimensions of [150 × 160 × 160] to create a dataset with uniform dimensions, and avoid memory allocation and computational overload errors. Every DICOM image volume from the scanner had a respective rescale slope [header 0028,1053] and intercept [header 0028, 1052] header values that resulted from the differing maximum and minimum intensities inherently present in the data. The image volumes look the same with respect to intensity, density, and overall image texture by creating uniform maximum and minimum intensities across all volumes captured by the scanner, despite a different amount of radioactivity (cell number) present amongst the images. To reverse the effect of the scanner's rescale factor and slope on the pixel intensities of the images, a rescale equation was formulated to reverse the effect of the inherent rescaling properties of the data by the scanner, thus revealing the true linear pattern of increasing radioactivity that is apparent with increasing total pixel sum (and increasing cell number). The rescale equation (Equation 1) is formatted as follows:

$$M_x = (M_r * R_s) + (R_i) \quad \text{(Equation 1)}$$

The rescale equation (Equation 1) generated a new pixel matrix, M_x, by multiplying the raw pixel matrix values, M_r, by the rescale slope from the DICOM header, R_s, and adding the product of these values to the rescale intercept from the DICOM header, R_i. This rescale equation was applied to all image volumes as a pre-processing step before the images and their corresponding radioactivity values were input to the CNN for training.

3D CNN architecture, training and cell number classification

The Convolutional Neural Network (CNN) was engineered using theKeras and TensorFlow libraries within Python (Wilmington, DE). The CNN algorithm consisted of a feature extraction, flattening layer, and regressing layer for complete analysis of the 3D PET data of input shape [150, 180, 180] after preprocessing and rescale by the aforementioned rescale and resampling steps (Figure 1). As shown in Figure 1, the feature extraction process includes three convolutional layers and three average pooling layers. The convolutional layer extracts the local features of the initial input, and the average pooling layer increases translational invariances of the network and reduces the required training parameters. The kernel size of each convolutional layer is [3, 3, 3], and the pooling size of each average pooling layer is two. The shape of the output of the last pooling layer is [17, 20, 20]. The flatten layer reshapes the 3D array to a 1D array to feed the feature into a fully connected layer with sixteen neurons. Then the regressor is fed the integrated information to make the final predictions. In this structure, the learning rate used was $\alpha = 10^{-5}$, the batch size was four, and the Adam optimizer was applied under 200 iterations. Loss was estimated using mean absolute error values from the previously established gradient descent loss algorithm.[34] A checkpoint algorithm function with an increased learning rate $\alpha = 10^{-4}$ was used in the case of no improvement in loss after 7 sequential iterations. Our dataset consists of 104 3D images with shape of [150, 180, 180]. A 5-fold cross-validation was applied to randomly divide 104 samples in five subgroups. All machine learning and deep learning models used in the study were trained using 4-folds as training set. The other remaining fold served as the test set for the evaluation.

Post-processing steps converted the predicted radioactivity value from the PET image sequence into an estimated cell number value depending on the cell line. A rearranged half-life equation for the ^{18}F radioactive isotope used the predicted radioactivity value and converted it to an initial radioactivity value, R_0, by considering the amount of time passed between the time of sample labeling and scanning, t, and the half-life for the ^{18}F isotope $t_{1/2}$ of 109.7 min. The equation below describes the relationship (Equation 2):

$$R_p = R_0\left(\frac{1}{2}\right)^{\frac{t}{t_{\frac{1}{2}}}} \quad \text{(Equation 2)}$$

Following conversion to R_0, the converted value was sorted into hard-coded classification bins based on the standardized ranges of radioactivity corresponding to different initial radioactivity values of varying cell numbers as was measured and recorded by the dose calibrator experiment. These classification bins were cell lines specific and consisted of 5 different ranges of radioactivity corresponding to 5 cell numbers: 3.125×10^5, 6.25×10^5, 1.25×10^{6}, 2.5×10^6, and 5×10^6 cells. As mentioned previously, the ß-TC-6 and MC38 cell lines had different classification bins due to their differing metabolic uptake of ^{18}F-FDG across all five potential cell number amounts. However, it is important to note that the CNN does predict cell numbers— the bins convert the value predicted by the CNN into a usable metric for inference on input data. Users can apply the algorithm to predict cell number if the standard ranges of radioactive uptake are known for a particular cell line without the classification bins. A schematic overview of the DL model with preprocessing, 3D CNN and post-processing cell number classification steps is shown in Figure 1.

2D CNN architecture

For comparison with the 3D CNN, a 2D CNN was built using Keras and TensorFlow libraries in Python. For the 2D CNN model, we divided the 3D images into 150 slices (channels) of 2D images with shape of (180, 180). As shown in Figure 2A, the feature extraction process includes four convolutional layers and four max pooling layers. The convolutional layer extracts the local features of the initial input, and the average pooling layer increases translational invariances of the network and reduces the required training parameters. The kernel size of each convolutional layer is (3, 3), and the pooling size of each max pooling layer is two. The shape of the output of the last pooling layer is (9, 9). The flatten layer reshapes 2D array to 1D array to feed the feature into a fully connected layer with sixteen neurons. Then the regressor is fed the integrated information to make the final predictions. In this structure, the learning rate is $\alpha = 10^{-4}$, the batch size is four, and the Adam optimizer is applied under 200 iterations.

The output predictions from the 2D CNN did not undergo subsequent post-processing steps of time-correction of the predicted radioactivity values and sorting into one of the classification bins to estimate cell number in the images. The 2D CNN was used solely for comparison with the 3D CNN and GBDT algorithms in the study.

Gradient-boosted decision tree architecture

As a basic machine learning algorithm, gradient boosting decision tree (GBDT) acts as an ensemble of decision tree with a flowchart-like structure that performs both classification and regression tasks. A GBDT predictor usually displays strong predictive power, which is robust to outliers. In the present work, we focus on the regression performance of the GBDT predictor and compare the algorithms performance to our primary 3D CNN algorithm employed throughout the study.

The GBDT model was set up through the following algorithm: Suppose we have a training set $\{(\mathbf{x}_i, y_i)|\mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}\}_{i=1}^n$. Initially, we built the first gradient boosting tree to fit training data, and the predicted label of Tree 1 is $p_1(\mathbf{x}_i)$. Then the residues were displayed as $r_2(\mathbf{x}_i) = y_i - p_1(\mathbf{x}_i)$. If Tree 1 does not fit the data perfectly, then $r_2(\mathbf{x}_i) \neq 0$. Therefore, we built Tree 2 to compensate for the shortcoming of the existing Tree 1. Tree 2 is designed to fit the dataset $\{(\mathbf{x}_i, r_2(\mathbf{x}_i))|\mathbf{x}_i \in \mathbb{R}^m, r_2(\mathbf{x}_i) \in \mathbb{R}\}_{i=1}^n$. Similarly, the predicted label of Tree 2 is $p_2(\mathbf{x}_i)$ and the residues can be showed as $r_3(\mathbf{x}_i) = r_2(\mathbf{x}_i) - p_2(\mathbf{x}_i) = y_i - p_1(\mathbf{x}_i) - p_2(\mathbf{x}_i)$, which leads the predicted label for initial training set $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ for by combining Tree 1 and Tree 2 to be $p_1(\mathbf{x}_i) + p_2(\mathbf{x}_i)$. We repeated the same procedure until $r_N(\mathbf{x}_i) \rightarrow 0$. Hence, N consecutive Trees are built as illustrated in Figure 2B. In general, the predictions of N Trees (Equation 3.1) are defined as

$$\widehat{y}_N(\mathbf{x}) = \sum_{j=1}^{N} p_j(\mathbf{x}) \quad \text{(Equation 3.1)}$$

The general loss function of Tree N (Equation 3.2) is defined as

$$L_N = \sum_{i=1}^{n} l(\widehat{y}_N(\mathbf{x}_i), y_i) \quad \text{(Equation 3.2)}$$

where the square loss (Equation 3.3)

$$l(\widehat{y}_N(\mathbf{x}_i), y_i) = \frac{(\widehat{y}_N(\mathbf{x}_i) - y_i)^2}{2} \quad \text{(Equation 3.3)}$$

is equipped. Then, the general loss function L can be minimized along the gradient direction (Equation 3.4)

$$-\frac{\partial L_N}{\partial p_N(\mathbf{x}_i)} = y_i - \widehat{y}_N(\mathbf{x}_i) = r_{N+1}(\mathbf{x}_i) \quad \text{(Equation 3.4)}$$

For the GBDT model, we first flatten the 3D images as a 1D array, which makes the feature size to be [# slices, z, x, y] = [104, 150, 180, 180]. Next, the initial feature of the training set was fed into the aforementioned GBDT regressor for the prediction. Here, our model consists of a total of 30,000 trees with a 0.01 learning rate. The predicted radioactivity outputs of the GBDT model were used solely for comparison with the 3D CNN and did not undergo further post-processing for time-correction and cell number classification.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data are presented as mean ± SD. Three metrics were used in this work to evaluate the performance of different machine learning and deep learning regression models: root-mean-square deviation (RMSE), mean absolute error (MAE), and Pearson correlation coefficient (PCC). Statistical comparisons between two groups were evaluated by Student *t* test and corrected by one-way ANOVA for multiple comparisons using the SPSS statistical software (IBM, Armonk, NY). Here, the *scikit-learn* package in Python was applied to calculate the RMSE, MAE, and PCC of each model. A similar K-Fold CV approach was used to split the dataset of 150 image volumes into five total sets with four sets used for training and one test for testing validation. For analysis of Intraclass Correlation Coefficient between model predicted measures of radioactivity, true calibrator measured doses and corresponding cell numbers, the F-test was utilized. A value of $p < 0.05$ was considered to indicate statistical significance.