Received: 12 January 2021 | Revised: 30 May 2021 | Accepted: 8 June 2021

DOI: 10.1002/mp.15063

RESEARCH ARTICLE

MEDICAL PHYSICS

Fast dynamic brain PET imaging using stochastic variational prediction for recurrent frame generation

Amirhossein Sanaat[1] | **Ehsan Mirsadeghi**[2] | **Behrooz Razeghi**[3,4] | **Nathalie Ginovart**[5,6] | **Habib Zaidi**[1,7,8,9,10]

[1]Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland

[2]Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran

[3]Department of Computer Sciences, University of Geneva, Geneva, Switzerland

[4]School of Engineering and Applied Sciences, Harvard University, Boston, USA

[5]Department of Psychiatry, Geneva University, Geneva, Switzerland

[6]Department of Basic Neurosciences, Geneva University, Geneva, Switzerland

[7]Geneva University Neurocenter, Geneva University, Geneva, Switzerland

[8]Department of Nuclear Medicine and Molecular Imaging, University of Groningen, Groningen, Netherlands

[9]University Medical Center, Groningen, Netherlands

[10]Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark

Correspondence
Habib Zaidi, Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland.
Email: habib.zaidi@hcuge.ch

Funding information
Swiss National Science Foundation, Grant/Award Number: 320030_176052 and 31003A_179373

Abstract

Purpose: We assess the performance of a recurrent frame generation algorithm for prediction of late frames from initial frames in dynamic brain PET imaging.

Methods: Clinical dynamic ^{18}F-DOPA brain PET/CT studies of 46 subjects with ten folds cross-validation were retrospectively employed. A novel stochastic adversarial video prediction model was implemented to predict the last 13 frames (25–90 minutes) from the initial 13 frames (0–25 minutes). The quantitative analysis of the predicted dynamic PET frames was performed for the test and validation dataset using established metrics.

Results: The predicted dynamic images demonstrated that the model is capable of predicting the trend of change in time-varying tracer biodistribution. The Bland-Altman plots reported the lowest tracer uptake bias (−0.04) for the putamen region and the smallest variance (95% CI: −0.38, +0.14) for the cerebellum. The region-wise Patlak graphical analysis in the caudate and putamen regions for eight subjects from the test and validation dataset showed that the average bias for K_i and distribution volume was 4.3%, 5.1% and 4.4%, 4.2%, (P-value <0.05), respectively.

Conclusion: We have developed a novel deep learning approach for fast dynamic brain PET imaging capable of generating the last 65 minutes time frames from the initial 25 minutes frames, thus enabling significant reduction in scanning time.

KEYWORDS
brain imaging, deep learning, dynamic imaging, PET, recurrent neural network

1 | INTRODUCTION

Positron emission tomography (PET) has ingrained wide clinical acceptance, particularly for its role in monitoring events at the molecular, and cellular levels in a number of diseases including neurological disorders.[1,2] In clinical setting, PET image interpretation is commonly carried out through visual analysis by comparing patient's image textures to typical patterns associated with various brain disorders. This is accompanied

by quantitative analysis through either comparison to disease-free templates reflecting normal tracer-specific biodistribution or extraction of image-derived PET metrics providing clinically relevant physiological parameters.[3]

Following the injection of a positron-emitting tracer to the patient, two data acquisition modes can be performed, namely static and dynamic scanning.[4,5] When the radiotracer distribution is relatively stable or quickly reaches an equilibrium state, a static scan is usually acquired. PET data acquisition using this scanning mode is commonly employed for tracers with virtually constant or slowly varying activity concentration over time within the target tissue, and is commonly acquired in a single frame. In clinical setting, routine [^{18}F]-Fluoro-2-deoxy-2-D-glucose (FDG) brain PET scanning is usually performed about 30 minutes post-injection to acquire a static scan. Dynamic brain PET imaging is commonly performed in research protocols to assess the time course of radiotracer uptake. In dynamic imaging performed to calculate blood flow or receptor occupancy, a sequence of dynamic frames is usually defined encompassing the whole duration of the PET scan. This imaging methodology commonly requires the extraction of arterial/venous samples to calculate the input function needed for tracer kinetic modeling to estimate relevant quantitative parameters,[6,7] although simplified blood sampling-free approaches have also emerged as an alternative approach.[8]

Dynamic imaging has been used in clinical routine for decades in nuclear medicine, particularly in planar imaging, single-photon emission computed tomography (SPECT) and, PET imaging.[9,10] Numerous studies using dynamic imaging in planar imaging, and multiple dynamic procedures continue to be employed.[11,12] For instance, bone scintigraphy, which evaluates the spread of active bone formation in the skeleton related to malignant/benign disease, is one of the most common nuclear medicine imaging procedures performed dynamically immediately after radiotracer injection to estimate perfusion.[13] Dynamic SPECT is another technique that emerged in the 1990s, mainly on rotating gamma cameras.[14,15] Dynamic and gated SPECT imaging has developed significantly more recently as it became more practical on dedicated solid-state cardiac cameras.[16] Dynamic cardiac PET imaging gained attention as a valuable technique in the clinic.[17] Another application of dynamic PET imaging is in clinical oncology, where compartmental modelling can potentially improve both tumor characterization, and treatment response monitoring.[6,18] Conventional image reconstruction of dynamic PET data based on independent handling of single frames has limited performance particularly in the initial dynamic frames owing to the low statistics.[7,19] Cui *et al*.[20] proposed a reconstruction method for dynamic PET imaging based on a stacked sparse auto-encoder. The dynamic reconstruction problem was expressed in a deep learning representation where the encoding layers extract the prototype features, such as edges, whereas the decoding layers generate the reconstructed images through a combination of these features. Hashimoto *et al*.[21] introduced a method to denoise dynamic PET images based on a deep image prior to enable unsupervised denoising without pretraining of datasets. A number of denoising strategies were applied to ultra-fast/low-dose whole body PET images.[22] Furthermore, a tracer-specific deep denoising autoencoder (DAE)-based approach was developed by Klyuzhin et al. that reduces the voxel-level noise in simulated dynamic [^{11}C]-raclopride brain PET images.[23] Rubinstein et al. developed a framework to detect prostate cancer using an unsupervised learning method for early detection, and localization of malignant lesions.[24] Their recommended method is able to extract features, including statistical, kinetic biological, and deep features from 4D imaging data through learning using a deep stacked convolutional auto-encoder.[24]

Although a number of applications require dynamic PET imaging, the long acquisition time limited the applicability of this methodology in clinical setting, especially for elderly and pediatric patients. Various strategies attempted to reduce the overall dynamic PET acquisition time in PET studies without using deep learning.[25,26] For instance, Torizuka *et al*. showed that a fast dynamic PET scan (30 minutes) of patients with lung cancer can lead to similar interpretation and compartment modeling results as regular 60 minutes dynamic PET scans.[27] Furthermore, Monden *et al*.[28] demonstrated that ^{18}F-FDG brain PET scanning time can be reduced from 60 minutes to 40 minutes without significant impact on the estimation of kinetic parameters (K_3 and K_i). The same approach was followed by Visser *et al*. to reduce dynamic scanning time from 50 minutes to 30 minutes in non-small cell lung carcinoma (NSCLC) patients.[29] More recently, Samimi *et al*. showed that a fast 5 minutes dynamic scan, accompanied with a 3 minutes static scan acquired 60 minutes post-injection enables to estimate accurately standardized uptake value (SUV) and 2-tissue-compartmental model parameters.[30] To the best of our knowledge, there is a lack of works using deep learning to decrease the acquisition time in dynamic PET imaging or to generate late dynamic frames from initial ones. In this work, we used a novel recurrent frame generation model, referred to as Stochastic Adversarial Video Prediction (SAVP), to reduce the acquisition time by estimating half of the late frames of dynamic ^{18}F-DOPA PET images from its first half frames. The SAVP model is mostly used for video prediction and is able to model the trend of time-varying activity distributions. To the best of our knowledge, this is the first work reporting on the feasibility of an approach enabling to reduce the scan time from 90 minutes to 25 minutes, thus making the scanning procedure more comfortable.

The main objective of this study was to decrease the scanning time while preserving the outcome by achieving the lowest bias. Shortening PET scanning time increases scanner throughput, and enhances patient comfort, thus reducing unwanted motion especially in elderly and pediatric patients.

2 | MATERIALS AND METHODS

2.1 | Brain PET/CT data acquisition

The present study was conducted on ^{18}F-DOPA dynamic brain PET/CT studies collected between October 2017, and December 2019 at Geneva University Hospital. ^{18}F-DOPA (DOPAVIEW®) was obtained from Advanced Accelerator Applications S.A. (Geneva, Switzerland). The database consisted of 46 subjects (27 males and 19 females; age = 255± yrs) including 12 healthy controls, 25 subjects with cannabis use disorder (CUD), and 9 patients with Internet gaming disorder (IGD). A tenfold cross-validation scheme was applied to prevent overfitting and selection bias and to check the generalizability of our model to new datasets. The demographic information of the subjects is summarized in Table 1. The study protocol was approved by the institution's ethics committee, and all patients gave written informed content. PET/CT acquisitions were performed on a Biograph mCT scanner (Siemens Healthcare, Erlangen, Germany). A low-dose CT scan (120 kVp, 20 mAs) was performed for PET attenuation correction. The patients underwent a brain PET scan lasting 90-minutes after injection of 190 ± 10 MBq of ^{18}F-DOPA. PET data were acquired in list-mode format and then binned into sinograms, and reconstructed into 26 different dynamic frames (2 × 30 seconds, 4 × 60 seconds, 3 × 120 seconds, 3 × 180 seconds, 14 × 300 seconds) as per the requirements of the research protocol using the e7 tool (an offline reconstruction toolkit provided by Siemens Healthineers). PET images were reconstructed into a 200 × 200 × 109 image matrix (2.03 × 2.2 × 2.03 mm^3 voxel size) using an ordinary Poisson-ordered subsets expectation maximization (OP-OSEM) algorithm (5 iterations, 21 subsets) considering time-of-flight information. PET images underwent post-reconstruction Gaussian filtering with 2 mm FWHM.

To make the model more sensitive and accurate and to reduce the computational time, we cropped the entire brain volume into two portions, each containing 30 slices (100 × 100 × 30 voxels), and trained the model twice for each section separately. The first brain portion was selected to contain the striatum consisting of the caudate nucleus and putamen, while the second portion was more caudal to contain the cerebellum, a region with negligible level of dopaminergic innervation, used as reference region to estimate nonspecific binding.

TABLE 1 Demographic characteristics of the subjects included in this study

	Training	Test	Validation
Number	35	8	3
Male/Female	27/19		
Age (Mean ±SD)	24±5		
Weight (Mean ±SD)	67±13		
Indication/Diagnosis	12 healthy controls 25 patients with Cannabis use disorder 9 patients with Internet gaming disorder		

2.2 | Learning model

We utilized the SAVP model proposed in Ref. [31] to generate raw pixels of future dynamic frames, given a sequence of initial frames. Due to the ambiguous nature of dynamic frames prediction, stochastic models outperform deterministic models and loss functions. As a result, recent research focused on Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) for image prediction, and generation problems.[32-40]

VAEs can explicitly model the underlying stochasticity. However, the prediction distribution in the VAE model is factored over pixels, thus leading to pixel-wise mean square error (MSE) loss, which may produce blurred, and unrealistic images. The GAN model overcomes this limitation; however, adversarial loss functions are difficult to tune, and are prone to mode collapse. To address these challenges, the VAE-GAN model proposed in Ref. [41] combines VAE-based latent variable model with an adversarial lose to improve the diversity of generated frames, thus enabling to produce realistic frames. Inspired by the above reference, the authors in Ref. [31] extended the results for SVP to predict diverse realistic frames, given the initial frames.

Based on this introduction, we now address our learning model. To this end, we first review the core idea of VAE, GAN, and Hybrid VAE-GANs.

2.2.1 | Variational Autoencoder (VAE)

The VAE model performs inference and generation by introducing a latent vector $\mathbf{Z} \in \mathcal{Z}$, in which it assumes that this latent vector follows a prior distribution $Q_{\mathbf{Z}}$.[42] Let $E_{\phi}(\mathbf{X}) \triangleq P_{\phi}(\mathbf{Z}|\mathbf{X})$ denotes an encoder parametrized by the output of a deep neural network with parameters ϕ. Let $G_{\theta}(\mathbf{Z}) \triangleq P_{\theta}(\mathbf{X}|\mathbf{Z})$ denotes a generator (decoder) parametrized by the output of the deep neural network with parameters θ. Let $P_{\mathrm{D}}(\mathbf{X})$ denotes an empirical data distribution. In this case,

$P_{\phi}(\mathbf{X},\mathbf{Z})$ denotes our joint inference data distribution, and $P_{\phi}(\mathbf{Z}) = \mathbb{E}_{P_D(\boldsymbol{X})}\left[P_{\phi}(\mathbf{Z}|\mathbf{X})\right]$ denotes the marginal posterior distribution over latentace $\mathcal{Z}$. Finally, let $P_{\theta}(\mathbf{X}) = \mathbb{E}_{Q_{\mathbf{Z}}}\left[P_{\theta}(\mathbf{X}|\mathbf{Z})\right]$ denotes the generated data distribution using the encoder $E_{\phi}(\mathbf{X})$, decoder $G_{\theta}(\mathbf{Z})$, and latent space prior $\mathrm{Q}_{\mathbf{Z}}$.

The β-VAE[43] relaxed the regularization term by introducing a Lagrangian multiplier. Hence, averaging over the empirical data distribution, the Lagrangian functional is defined as:

$$\mathcal{L}_{\beta-\mathrm{VAE}}(E_{\phi},G_{\theta}) = \mathbb{E}_{P_D(\mathbf{X})}\left[\mathbb{E}_{P_{\phi}(\mathbf{Z}|\mathbf{X})}\left[\log P_{\theta}(\mathbf{X}|\mathbf{Z})\right] - \beta \mathrm{D}_{KL}\left(P_{\phi}(\mathbf{Z}|\mathbf{X}=\mathbf{x})\|\mathrm{Q}_{\mathbf{Z}}\right)\right] \quad (1)$$

During training, we sample latent codes from two distributions: (a) the prior Gaussian distribution and (b) a posterior distribution approximated by a learned encoder Eφ. In order to have a stable gradient with respect to the encoder, a reparametrization trick was used to sample from the learned posterior distribution.[42] During testing, we sampled the latent codes from a fixed unit Gaussian prior, N (0, 1), as our prior distribution Q(z). Since we assume Gaussian distribution, L1 loss was reduced to mean square error.

2.2.2 | Generative Adversarial Network (GAN)

The GAN[44] consists of two networks: the generator network $\mathrm{G}_{\theta}(\mathbf{Z})$ maps a latent vector $\mathbf{Z} \in \mathcal{Z}$ to data space $\hat{\mathbf{X}} \in \mathcal{X}$ and a discriminator network $\mathrm{D}_{\omega}(\mathbf{X})$ to distinguish whether the samples are coming from the empirical data distribution $\mathrm{P}_{\mathrm{D}}(\mathbf{X})$ or generated distribution $\mathrm{P}_{\theta}(\mathbf{X})$. Hence, the discriminator network $\mathrm{D}_{\omega}(\mathbf{X})$ acts like a classifier that assigns probability $y = \mathrm{D}_{\omega}(\mathbf{X}) \in [0,1]$ if $\mathbf{X}$ is an actual training sample, and probability $1-y$ if $\mathbf{X}$ is generated by the model through $\hat{\mathbf{X}} = G_{\theta}(\mathbf{Z})$ with $\mathbf{Z} \sim Q_{\mathbf{Z}}$. The original GAN[44] problem aims to maximize/minimize the binary cross entropy (an adversarial game):

$$\mathcal{L}_{\mathrm{GAN}}(G_{\theta},\mathrm{D}_{\omega}) = \mathbb{E}_{\mathrm{P}_{\mathrm{D}}(\mathbf{X})}\left[\log \mathrm{D}_{\omega}(\mathbf{X})\right] + \mathbb{E}_{\mathrm{Q}_{\mathbf{Z}}}\left[\log\left(1-\mathrm{D}_{\omega}(\mathrm{G}_{\theta}(\mathbf{Z}))\right)\right] \quad (2)$$

with respect to Discriminator/Generator. Theoretically, when the adversarial process reaches the Nash equilibrium, the mini-max game attains its global optimum.

2.2.3 | Hybrid VAE-GANs

The VAE-GAN combines the VAE and GAN models to utilize the appealing properties of these generative models. By using the reparametrization trick, back-propagation was implemented in the encoder:

$$\mathcal{L}_1 = \mathbb{E}_{\mathrm{P}_{\mathrm{D}}(\mathbf{X})}\left[\mathbb{E}_{\mathrm{P}_{\phi}(\mathbf{Z}|\mathbf{X})}\left[\log \mathrm{P}_{\theta}(\mathbf{X}|\mathbf{Z})\right]\right] \quad (3)$$

A regularization term was considered to enable sampling from the prior at test time. This term promotes the estimated posterior to be as close as possible to the prior distribution:

$$\mathcal{L}_{\mathrm{prior}} = \mathbb{E}_{\mathrm{P}_{\mathrm{D}}(\mathbf{X})}\left[\mathrm{D}_{\mathrm{KL}}\left(\mathrm{P}_{\phi}(\mathbf{Z}|\mathbf{X}=\mathbf{x})\|\mathrm{Q}_{\mathbf{Z}}\right)\right] \quad (4)$$

Hence, the main goal of VAE is to minimize the summation of the above-mentioned functions:

$$\mathcal{L}_{\beta-\mathrm{VAE}}(E_{\phi},\mathrm{G}_{\theta}) = \lambda_1\mathcal{L}_1 + \beta\,\mathcal{L}_{\mathrm{prior}} \quad (5)$$

The original VAE-GAN paper[41] replaces the log-likelihood term $\mathcal{L}_{\mathrm{rec}}$ by a special similarity measure in the latent space which is trained through a GAN discriminator.

2.2.4 | Conditional VAE-GANs

The conditional VAE-GAN[45] uses both the latent code $\mathbf{Z}$ and the input image $\tilde{\mathbf{X}}$ to synthesize the desired output $\hat{\mathbf{X}}$. That is, Conditional VAE-GAN starts from a ground truth target image $\mathbf{X}$ and encodes it into the latent vector $\mathbf{Z}$. Next, the generator $G_{\theta}(\mathbf{Z},\tilde{\mathbf{X}})$ reconstructs $\hat{\mathbf{X}}$ using a sampled latent vector $\mathbf{Z}$ concatenated with the paired $\tilde{\mathbf{X}}$. Note that the conditional VAE-GAN can be viewed as a combination of conditional VAE and conditional GAN models. Although, GANs can suffer from the problem of mode collapse, especially in the conditional generation setting, conditional VAE-GAN mitigates this problem by introducing the prior distribution for latent space $\mathcal{Z}$.

2.3 | Learning algorithm

Our model is based on VAE-GAN and their conditional variants. We used VAE-GAN in a recurrent setting for prediction of late frames from initial frames in dynamic brain PET imaging. During the training phase (Algorithm 1), a recurrent generator network G_{θ} predicts new frames $\hat{\mathbf{x}}_t$, given the random latent code $\hat{\mathbf{z}}_{t-1}$ and the previous frame(s) $\tilde{\mathbf{x}}_{t-1}$. The generator is a convolutional Long Short Term Memory (LSTM). To simplify the notation, the input $\tilde{\mathbf{x}}_{t-1}$ denotes the ground truth frames or the previous predicted frames. The encoder E_{ϕ} denotes a deep network that is conditioned on the ground truth adjacent frames to encode them into a latent code $\mathbf{z}_t$. Our encoder is a feed-forward convolutional neural network. During the testing phase, a recurrent generator network G_{θ} projects the initial frame(s) and a sequence of latent random codes $\hat{\mathbf{z}}_{0:T-1}$ to the predicted frames $\hat{\mathbf{x}}_{0:T-1}$. The block-diagram of the model is

Algorithm 1 Training the stochastic variational prediction

$\boldsymbol{\varphi, \theta, \omega} \leftarrow$ **network initialization**
Repeat:
  $\mathbf{x}_{1:T} \leftarrow$ **random batch from the training set**
  $\mathbf{x}_{t-1}, \mathbf{x}_t \leftarrow$ **Reference frame** $(t-1, t)$ **as input**
  $z_{t-1} \leftarrow E_{\varphi}(\mathbf{x}_{t-1}, \mathbf{x}_t)$
  $\mathcal{L}_{prior} \leftarrow \left[D_{KL}\left(P_{\phi}(Z|X=x) \parallel Q_Z\right)\right]$
  $\hat{\mathbf{x}}_t \leftarrow G_{\theta}\ (z_{t-1}, \tilde{\mathbf{x}}_{t-1})$
  $\mathcal{L}_{GAN} \leftarrow D_{\omega}(\hat{\mathbf{x}}_t, \mathbf{x}_{1:T})$
  $\mathcal{L}_1 \leftarrow \mathrm{D\_VAE}(\hat{\mathbf{x}}_t, \mathbf{x}_{1:T})$
  $\mathcal{L}_{\beta-\mathrm{VAE}} \leftarrow \lambda_1 \mathcal{L}_1 + \beta\, \mathcal{L}_{prior}$
  $\mathcal{L}_{\mathrm{VAE-GAN}} \leftarrow \gamma\, \mathcal{L}_{GAN} + \mathcal{L}_{\beta-\mathrm{VAE}}$
  $\boldsymbol{\varphi, \theta, \omega} \leftarrow$ **update using backprob**
Until convergence
Return $\boldsymbol{\varphi, \theta, \omega}$

depicted in Figure 1. The visual space discriminators D_{GAN} and D_{VAE} denote the corresponding discriminator D_{ω} for the GAN and VAE setups respectively. Finally, $\mathscr{L}_1$ denotes the ℓ_1 loss between the output and the ground truth frame(s). This loss encourages the output of the generator to match the input and stabilizes the training.

2.4 | Training details

The model was trained using Adam[46] for 650000 iterations, linearly decaying the learning rate to 0 for the last 20000 iterations. GAN-based variants used an optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$, a learning rate of 0.0002, and a batch size of 26. The VAE models used an optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, a learning rate of 0.001, and a batch size of 26. For GAN-based variants, a $\lambda_1 = 100$ was used. This hyper-parameter was empirically selected by computing similarity metrics on the validation set. During the training, the KL divergence weight is changed linearly from zero to the final value of 0.00001.

To ensure a balanced data distribution for train, validation and test datasets, we used a simple t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize high-dimensional data by assigning each data point a location in a two- or three-dimensional map. This enables to train and test models with a similar portion of data (normal and abnormal). For instance, we first entered the whole dataset (46 cases) to t-SNE model to categorize them into five major groups. In the next step, we collected the same portion from each group for training, validation, and testing. This procedure improves model robustness and generalizability when feeding with new data. The training and testing were performed on a personal computer running windows 10 equipped with NVIDIA Quadra K5200 graphics processing unit with 8 GB of memory. The main loss function that requires optimization is given in Equation 5 representing the sum of variational and adversarial loss. The plots of these loss functions are shown in Supplemental Figure 1.

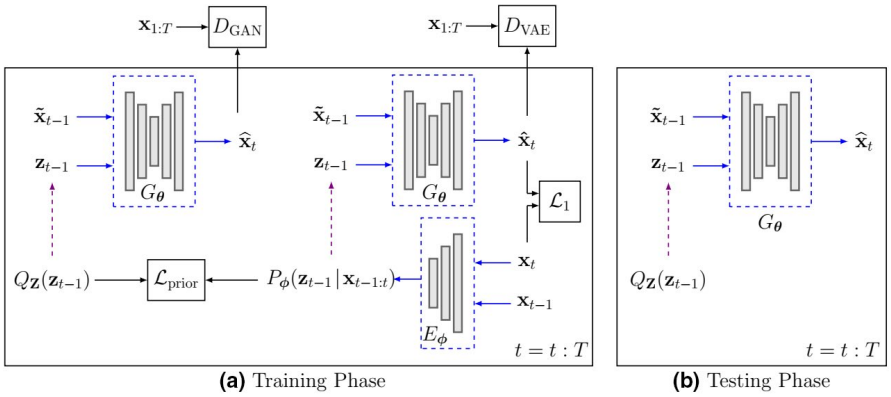

FIGURE 1 Block diagram of our conditional VAE-GAN model for prediction of late frames from initial frames in dynamic brain PET imaging: (a) training phase and (b) testing phase [Color figure can be viewed at wileyonlinelibrary.com]

2.5 | Data augmentation

To avoid over-fitting during the training and increase the size of the training dataset, two types of data augmentation methods, including translation and zooming, were randomly applied to the training dataset. Applying the above-mentioned rigid deformation to the dataset enabled the network to learn features that are invariant to these translations.

2.6 | Valuation strategy

The accuracy of the predicted second half of the frames sequence (frames 14–26) of dynamic images from the first half of the frames sequence (frames 1–13) was evaluated for each frame using three quantitative metrics, namely the root-mean-square-error (RMSE), the peak signal-to-noise-ratio (PSNR), and the structural similarity index measure (SSIM) (Equations 6-8, respectively).

$$\mathrm{RMSE}(\mathbf{x},\mathbf{y}) = \sqrt{\frac{\sum_{j=1}^{L}(x_j - y_j)^2}{L}} \tag{6}$$

$$\mathrm{PSNR}(\mathbf{x},\mathbf{y}) = 20 \times \log_{10}\left(\frac{\max_j(y_j)}{\sqrt{\mathrm{MSE}(\mathbf{x},\mathbf{y})}}\right) \tag{7}$$

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{(2m_{\mathbf{x}}m_{\mathbf{y}} + c_1)(2\sigma_{\mathbf{xy}} + c_2)}{\left(m_{\mathbf{x}}^2 + m_{\mathbf{y}}^2 + c_1\right)\left(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + c_2\right)} \tag{8}$$

In Equation (6), L is the number of voxels in the brain region, $\mathbf{x}$ represents the reference frame, and $\mathbf{y}$ the predicted frame. In Equation (7) $\max_j(y_j)$ shows the maximum intensity of $y_j \in \mathbf{y}$, whereas MSE is the mean-squared-error. The $m_{\mathbf{x}}$, $m_{\mathbf{y}}$ in Equation (8) represent the mean value of the frames $\mathbf{x}$, and $\mathbf{y}$, respectively. $\sigma_{\mathbf{xy}}$ is the covariance of of $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{y}}$, which in turn denotes the variances of $\mathbf{x}$, and $\mathbf{y}$ images respectively. The constant coefficients c_1, and c_2 ($c_1 = 0.01$ and $c_2 = 0.02$) were applied to avoid a separation by very tiny values.

Region-based analysis was also performed to assess the agreement between predicted and reference frames in terms of SUV, defined as the ratio between activity concentration (MBq/cc) in a defined region, and the injected activity (MBq) normalized to subject's weight (g).

Using the PMOD medical image analysis software (PMOD Technologies LLC, Switzerland) and the Hammers N30R83 brain atlas, five ROIs containing the caudate nucleus, putamen (left and right), and cerebellum were delineated on the reference PET frames. Since the PMOD software relies only an on FDG brain template for normalization, we replaced it with a ^{18}F-DOPA template available as part of the statistical parametric mapping package. This template is derived from 12 control subjects without evidence of nigrostriatal degeneration adjusted to Montreal Neurological Institute (MNI space).[47] Subsequently, the delineated ROIs were copied onto the predicted dynamic PET frames to quantify the SUV in each region, and for each time frame. To compare the biodistribution in the predicted, and reference dynamic frames, region-wise time activity curves (TAC) were drawn.

^{18}F-DOPA PET parametric images were generated using Patlak graphical analysis, and the reference tissue model[48] implemented in PMOD software. Maps of the influx rate constant (K_i in minutes^{-1}) and distribution volume (V) were generated using the dynamic PET frames, and a TAC obtained over the reference region corresponding to the cerebellum.[49] This technique is appropriate when the tracer is irreversibly trapped in tissue. We hypothesized that the bias would be higher at later time frames compared to early time frames. Therefore, the bias on parameter estimates (K_i, and V) generated using the Patlak model was evaluated on the full (90 minutes), and a truncated number of dynamic PET frames (65 minutes).

The MedCalc software[50] was employed for the calculation of the pairwise t-test for statistical analysis of RMSE, SSIM, and PSNR between predicted dynamic frames and reference dynamic frames. The significance level was set at a P-value <0.05 for all comparisons.

3 | RESULTS

The predicted dynamic frames (14–26) provided almost similar appearance and texture with respect to reference frames (Figure 2). The predicted frames show a slight gradual deterioration from the initial predicted frame (frame number 14) to the last predicted frame (frame number 26). Visual inspection revealed that the predicted images exhibited a similar biodistribution pattern of radiotracer uptake compared to reference images. Line profiles drawn through coronal views of predicted and reference images are shown in Figure 3. This figure shows a single slice at the middle of the brain containing a cross section of the caudate and putamen from the predicted and reference dynamic images (26th time frame of a single slice). Therefore, each slice depicts different time frames, thus presenting the profile as a pattern of a TAC curve.

Figure 4 depicts the PSNR, SSIM, and RMSE calculated on the test dataset for each predicted frame from 14 to 26. Overall, the predicted images in the earlier frames (14–20) provide higher image quality, better noise properties, and higher quantitative accuracy than the last frames (21–26). Although all three metrics get worse gradually over time, their magnitude even in the final frames, is not out of range.

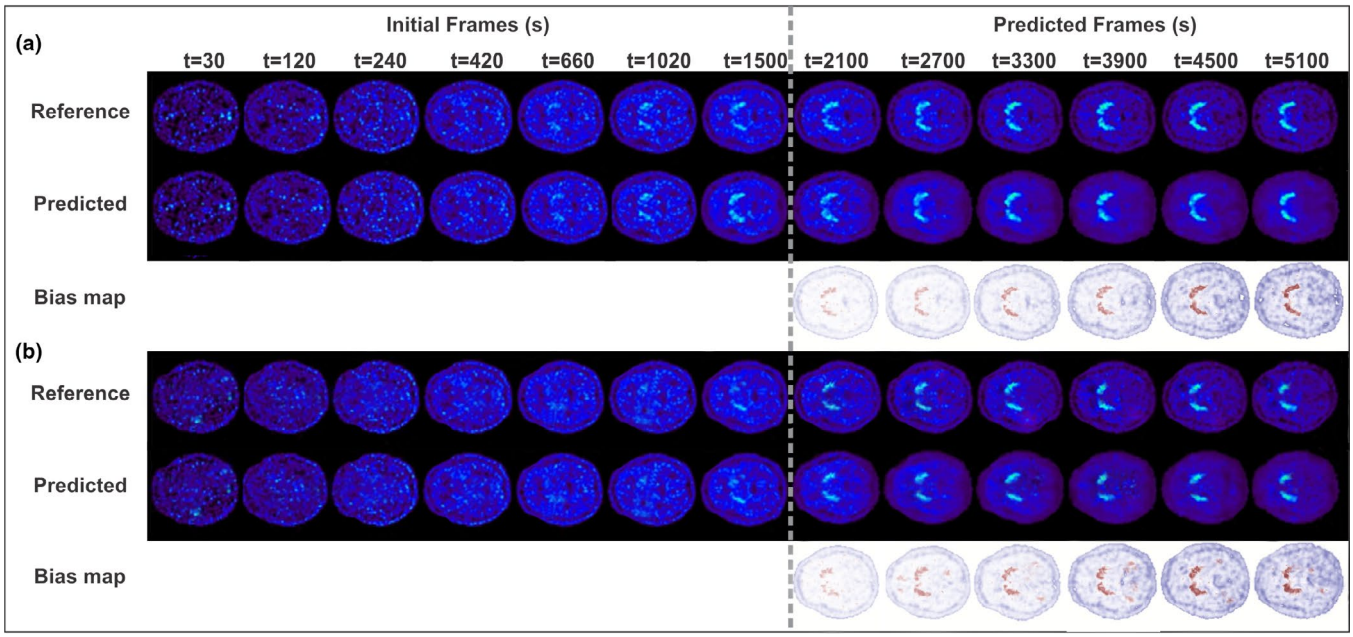

FIGURE 2 Comparison between reference and predicted frames for (a) a 26-year-old normal female subject and (b) a 27-year-old abnormal male subject. For presentation limitation, only odd frames are shown. The threshold for bias maps was kept constant (within ±5%) [Color figure can be viewed at wileyonlinelibrary.com]

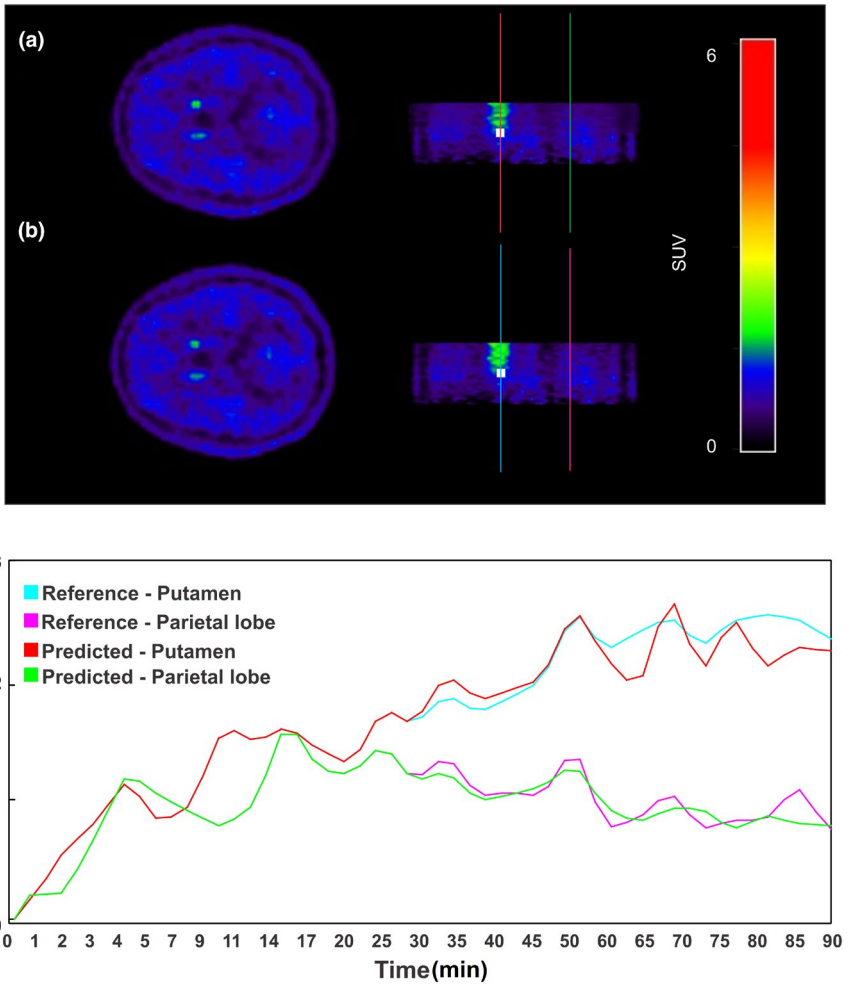

FIGURE 3 Comparison between predicted (a) and reference (b) time frames of a 26-year old female (top) along with line profiles drawn through the putamen and parietal lobe (bottom) of both images [Color figure can be viewed at wileyonlinelibrary.com]

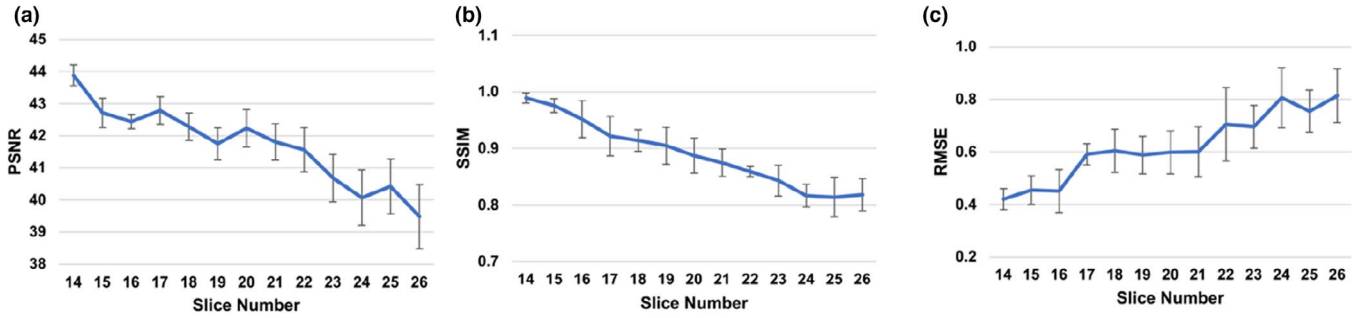

FIGURE 4 Image quality metrics comparing the predicted and reference last 13 dynamic frames for the test dataset. (a) PSNR: peak-signal-to-noise-ratio, (b) SSIM: structural similarity index measure, and (c) RMSE: root-mean-square-error [Color figure can be viewed at wileyonlinelibrary.com]

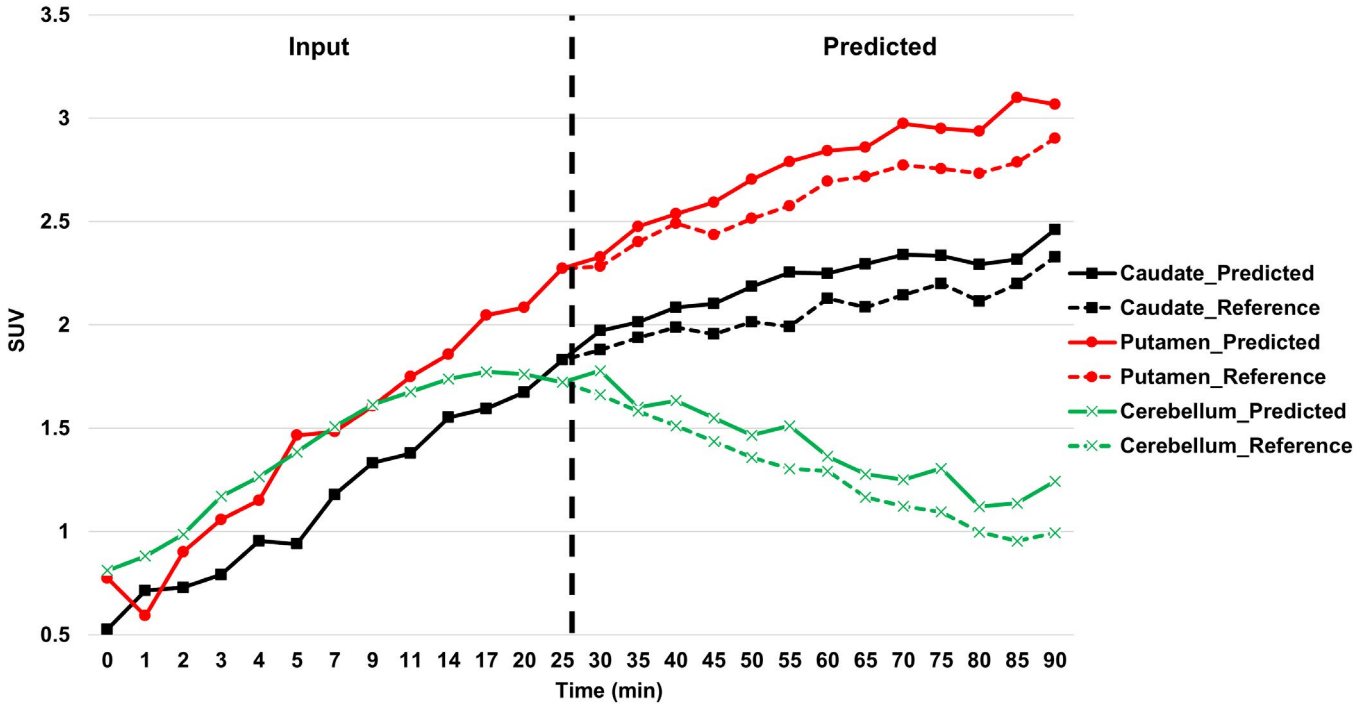

FIGURE 5 Comparison between reference and predicted time activity curves for the caudate, putamen, and cerebellum (average of left and right regions) for a 26-year old normal female subject. The black dotted line depicts the separation between the reference and the predicted frames [Color figure can be viewed at wileyonlinelibrary.com]

Figure 5 illustrates the TAC plots of predicted and reference images for three regions. The left side of the dashed line shows the input frames whereas the right side compares the reference and predicted frames.

The Bland-Altman plots, where each data point reflects the SUV_{mean} of each region for every time frame, confirmed the results obtained from parametric analysis (RMSE) where the lowest SUV bias is apparent in the initial frames whereas the late frames present the higher SUV bias (the frames'number of some data points 14–26 is mentioned) (Figure 6).

The plots show that the Putamen shows the lowest SUV bias (−0.04) whereas the smallest SUV variance (95% CI: −0.38, +0.14) was observed in the cerebellum. The SUV bias is extremely low for the initial frames, and the variance does not go further (95% CI:-0.38, +0.29), reflecting the good quantitative accuracy. Table 2 further confirms that the bias in parameter estimates (K_i, V) decreases for a representative number of subjects when a truncated number of dynamic PET frames (65 minutes) were used for Patlak analysis compared to using the full number of dynamic frames (90 minutes). To figure out how much our model was successful in improving kinetic parameter estimates, we calculate the kinetic parameters for only 0–25 minutes from reference frames and then compared the results with kinetic parameters derived from 0 to 90 minutes frames.

Figure 7 depicts the region-wise Patlak graphical analysis results comparing the estimated K_i from the predicted and reference dynamic frames for the left and right parts of the caudate and putamen for the test, and validation (Figure S2) datasets respectively. Figure 8 (and Figure S3) shows radar plots of the bias of the influx rate constant (K_i) and distribution volume (V) from

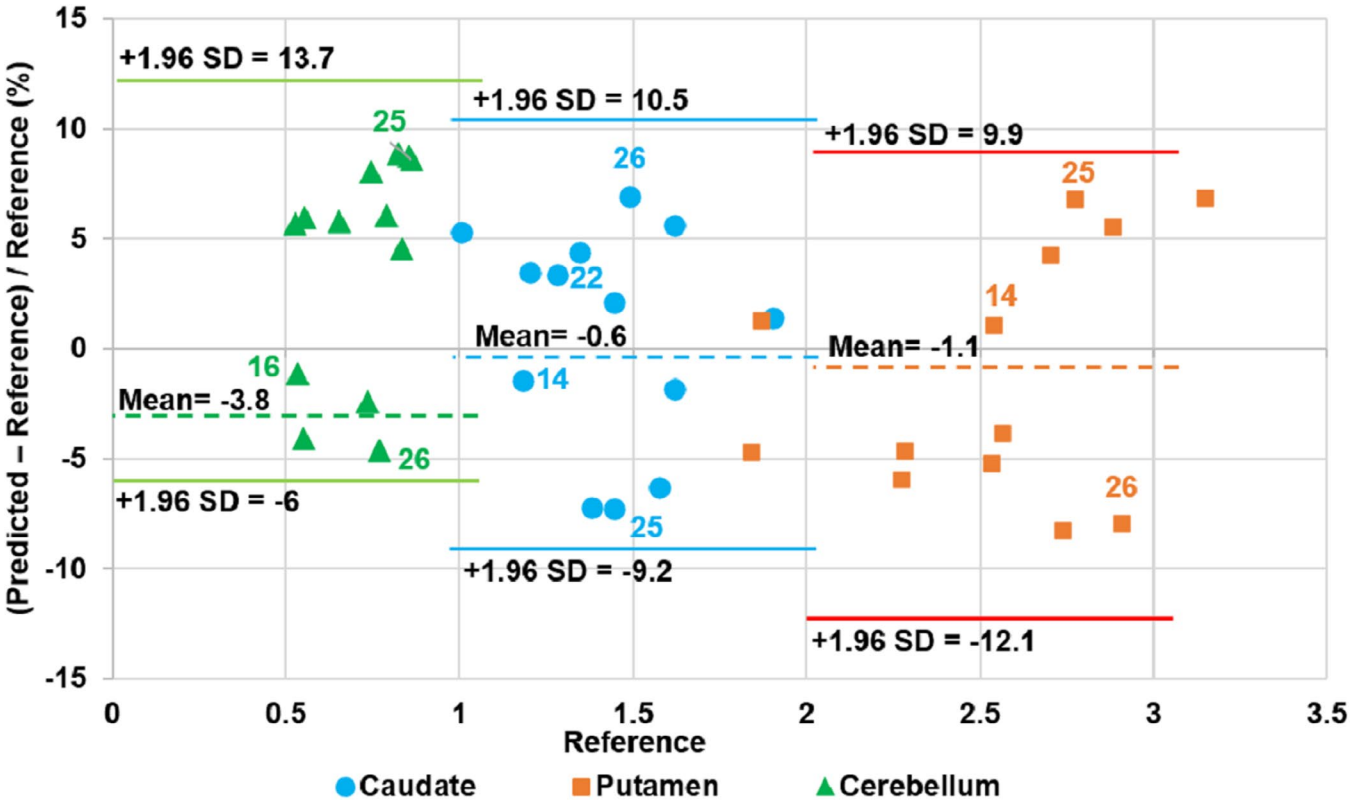

FIGURE 6 Bland & Altman plots of SUV differences in three brain regions (average of left and right) calculated for 14 predicted frames with respect to the reference frames in the test dataset. The dashed and solid lines denote the mean and 95% confidence interval (CI) of the SUV differences respectively. The colorful numbers shown near some data points show the frame numbers (from 14 to 26) [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Change in average influx rate constant (K_i) and distribution volume (V) resulting from graphical Patlak analysis of dynamic ^{18}F-DOPA brain PET studies for eight subjects from the test dataset when using the full (0–90 min predicted vs 0–90 min reference) and truncated (25–90 min predicted vs 25–90 min reference) PET data. The bias for the 0–25 min was calculated regarding full PET data (0–25 min reference vs 0–90 min reference)

Brain region	Average bias	0–25 min	25–90 min	0–90 min	P-value (25–90 min vs 0–90 min)
Caudate	Average K_i bias (%)	36.3 ± 12%	4.3 ± 4.5	5.1 ± 3.2	<0.05
	Average V bias (%)	33.7 ± 9.5%	3.3 ± 3.8	4.3 ± 2.2	<0.05
Putamen	Average K_i bias (%)	29.9 ± 7.6%	3.3 ± 2.7	4.4 ± 1.5	<0.05
	Average V bias (%)	39.4 ± 6.0%	3.8 ± 3.1	4.2 ± 2.8	<0.05

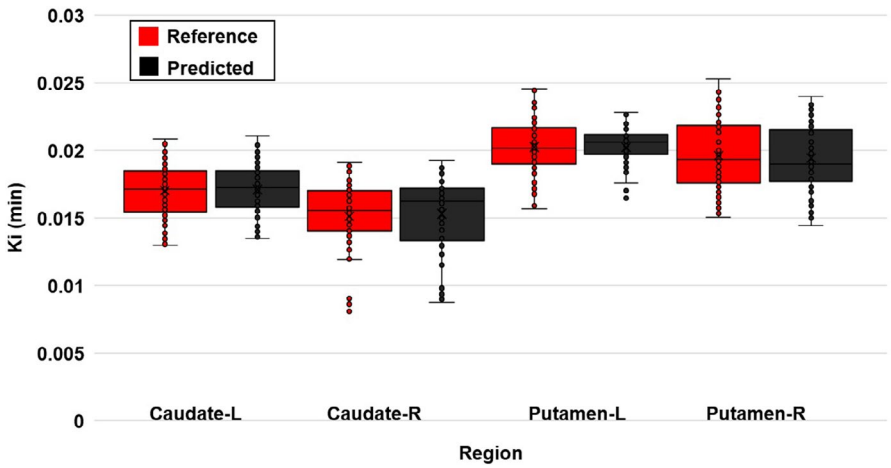

FIGURE 7 Comparison of the influx rate constant (K_i) between predicted and reference images for four brain regions, including the caudate and putamen (left and right) with the cerebellum used as reference region for the test datasets [Color figure can be viewed at wileyonlinelibrary.com]

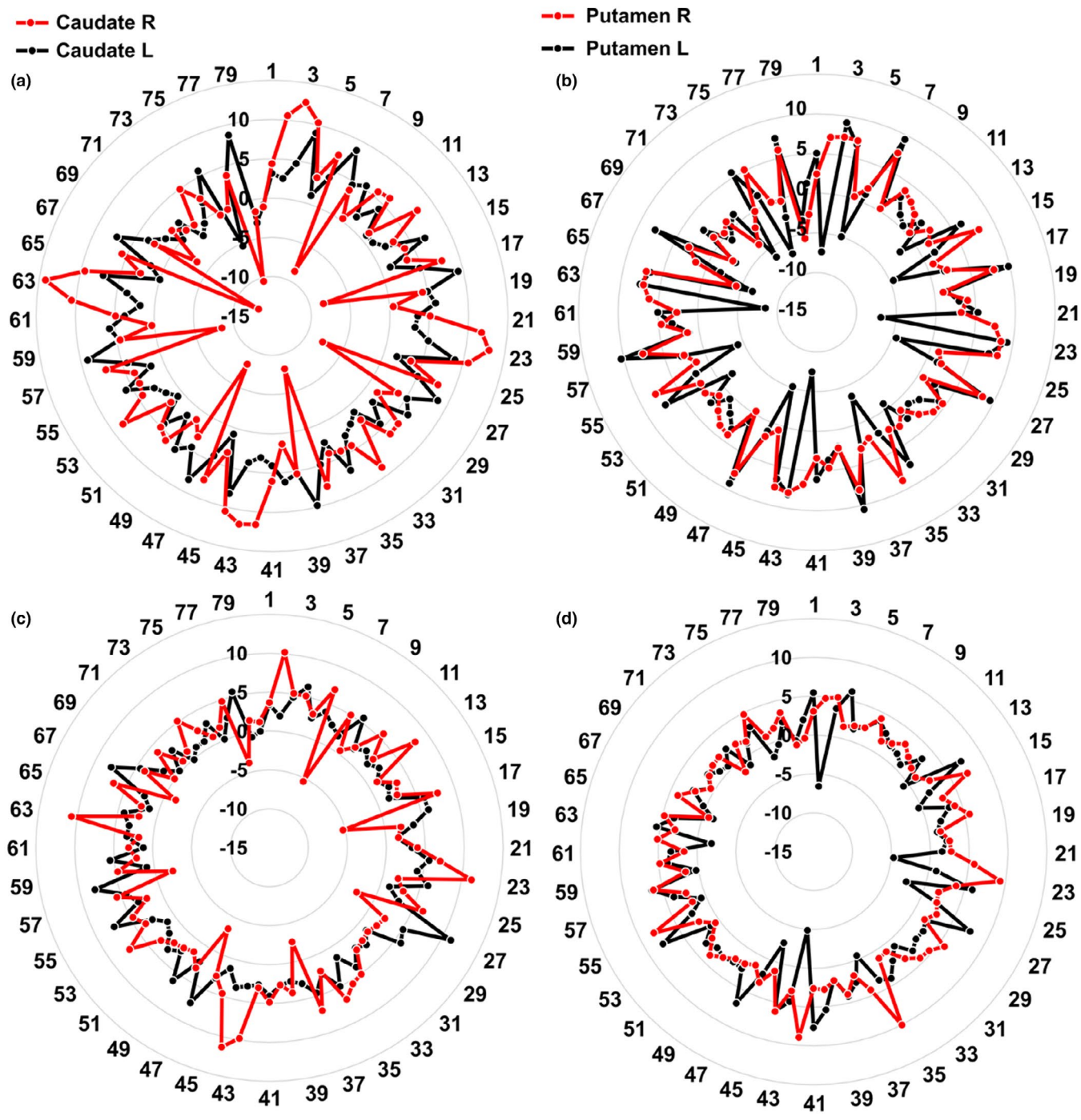

FIGURE 8 Radar plot of bias of the influx rate constant (K_i) (a and b) and distribution volume (V) (c and d) for the Caudate and Putamen, respectively, resulting from graphical Patlak analysis of dynamic ^{18}F-DOPA brain PET studies for subjects from the test dataset [Color figure can be viewed at wileyonlinelibrary.com]

the test and validation dataset respectively. The highest K_i and distribution volume biases for the test dataset are 7.14%, and 14.13% respectively.

4 | DISCUSSION

In this work, we aimed to generate last half frames (13 frames from 25 to 90 minutes) of dynamic brain PET images from the initial half frames (13 frames from 0 to 25 minutes). The goal is to synthesize full scan diagnostic quality dynamic ^{18}F-DOPA brain PET images from 27% of scanning time post-injection. The training of the neural network was performed using a SAVP model, considering initial 13 frames time series as input to predict the last 13 frames. The purposed deep learning model considers the initial time frames to capture the trend of the underlying biodistribution for

the latter frames. To the best of our knowledge, this is the first study considering this approach to decrease the scanning time of long dynamic PET imaging protocols using machine learning algorithms. The synthesized dynamic PET images predicted from the earlier half frames had comparable image quality and tracer uptake patterns, K_i and distribution volume values, and small variance relative to actual reference dynamic frames, especially for the first few predicted dynamic frames.

The RMSE calculated on synthesized images degrades from 0.35 for frame number 14 to 0.77 for frame number 26. The reason of this degradation in quantitative accuracy with time finds its root in the Markovian dependence of time frames, and Data Processing Inequality (DPI) concept. From one hand, the sequential frames form a Markov chain such that each predicted frame depends on the previous frame(s). Conversely, according to the DPI concept, we can only lose information through its processing. Let us consider the prediction process as a channel whose input are the previous frame(s), and its output is the next predicted frame. The DPI concept states that by passing data through a channel, the information cannot be increased. Therefore, the far frames degrade more than the near frames with subsequent increase of the RMSE. Moreover, the SSIM as a metric for calculation of similarity between the predicted and reference dynamic frames varies from 0.98 for the first predicted frame to 0.81 for the last frame.

The quantitative analysis of five brain regions in terms of TACs proved the effectiveness of our model in finding the optimal fitting curve for the radiotracer distribution. The quantitative evaluation of predicted images showed less than 7% bias for K_i, and less than 14% bias for V in most brain regions. Samimi *et al*. reported ~10% bias for kinetic parameters estimation when decreasing the dynamic scanning time to 5 minutes supplemented by a whole body PET scan 60 minutes postinjection.[30] A similar trend was also reported in two other studies where the bias in K_i estimates was ~5%[28] and 9%[27] when decreasing dynamic scanning time from 60 to 40 minutes and from 60 to 30 minutes respectively.

Although there is no evidence that CUD and IGD subjects have neurological abnormalities that could be reflected in differences in ^{18}F-DOPA uptake with respect to normal subjects, our aim was to build a heterogeneous dataset including subjects with and without brain disorders. Data augmentation techniques consisting of rotations and translation was applied to time series or dynamic frames to avoid overfitting and guarantee robust, and effective training. The Bland & Altman analysis showed low bias and variance in the three regional SUVmean values obtained from predicted dynamic PET images compared with reference images. The Bland & Altman plots further demonstrated the superior performance of the model in predicting the initial frames (e.g. frames 14–20) compared to the last frames (e.g. frames 21–26) resulting in SUV values comparable to the original images.

Dynamic PET images are commonly reconstructed successively using algorithms designed for static imaging. This approach is not optimal because the intrinsic temporal correlation between information present in dynamic series is not considered. A more thorough optimal approach would treat the whole dynamic data instead of considering them as a sequence of separate, and independent time frames. However, this process is time consuming.[19] In this work, since our proposed model uses previous images of a certain frame for prediction of that particular frame, it should learn to capture the noise model and biodistribution for a series of images, not a single image, and consider it for prediction of each new frame. In terms of computational time, the training took around 4 days for the dynamic 90 minutes ^{18}F-DOPA studies used in this work. Yet, the synthesis of a single dynamic PET image (after training) took only ~2 minutes.

One of the limitations of the present study was the limited sample size though it was almost tripled through data augmentation. Moreover, subject motion is likely to occur during the long PET scan, which might impair image quality at different time frames. Lastly, we did not train the model for the whole brain volume to reduce the computational load, and to avoid divergence of the model in regions where there is no sensible tracer uptake.

5 | CONCLUSION

We have introduced a novel application of variational autoencoder and generative adversarial network model for fast dynamic brain PET imaging which considers only the initial 25 minutes time frames to generate the last 65 minutes time frames of dynamic PET images using a deep learning approach for time series data. The model is promising, demonstrating reasonable performance considering the resulting good image quality, and relatively low PET quantification bias for the evaluated metrics.

AKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under grants SNRF 320030_176052 and 31003A_179373), by the Louis-Jeantet Foundation with contributions of the Clinical Research Center, University Hospital and Faculty of Medicine, Geneva.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are not available.

REFERENCES

1. Hooker JM, Carson RE. Human positron emission tomography neuroimaging. *Annu Rev Biomed Eng*. 2019;21:551–581.
2. Hope TA, Fayad ZA, Fowler KJ, et al. State of the Art PET/MRI: applications and limitations - Summary of the first ISMRM/SNMMI Co-Provided Workshop on PET/MRI. *J Nucl Med*. 2019;60(10):1340–1346.
3. Zaidi H, Karakatsanis N. Towards enhanced PET quantification in clinical oncology. *Br J Radiol*. 2018;91(1081):20170508.
4. Vallabhajosula S. *Molecular imaging: radiopharmaceuticals for PET and SPECT*. Berlin, Germany: Springer Science & Business Media; 2009.
5. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009;50(Suppl 1):11S–20S.
6. Rahmim A, Lodge MA, Karakatsanis NA, et al. Dynamic whole-body PET imaging: principles, potentials and applications. *Eur J Nucl Med Mol Imaging*. 2019;46(2):501–518.
7. Rahmim A, Tang J, Zaidi H. Four-dimensional (4D) image reconstruction strategies in dynamic PET: beyond conventional independent frame reconstruction. *Med Phys*. 2009;36(8):3654–3670.
8. Bentourkia M, Zaidi H. Tracer kinetic modeling in PET. *PET Clin*. 2007;2(2):267–277.
9. Madsen MT. Recent advances in SPECT imaging. *J Nucl Med*. 2007;48(4):661–673.
10. Rahmim A, Zaidi H. PET versus SPECT: strengths, limitations and challenges. *Nucl Med Commun*. 2008;29(3):193–207.
11. Sarikaya I, Sarikaya A. Current status of radionuclide renal cortical imaging in Pyelonephritis. *J Nucl Med Technol*. 2019;47(4):309–312.
12. Sanchez-Crespo A. Lung scintigraphy in the assessment of aerosol deposition and clearance. *Semin Nucl Med*. 2019;49(1):47–57.
13. Khawar A, Eppard E, Roesch F, et al. Biodistribution and post-therapy dosimetric analysis of [(177)Lu]Lu-DOTA(ZOL) in patients with osteoblastic metastases: first results. *EJNMMI Res*. 2019;9(1):102.
14. Gullberg GT, Reutter BW, Sitek A, Maltz JS, Budinger TF. Dynamic single photon emission computed tomography—basic principles and cardiac applications. *Phys Med Biol*. 2010;55(20):R111–R191.
15. Celler A, Farncombe T, Bever C, et al. Performance of the dynamic single photon emission computed tomography (dSPECT) method for decreasing or increasing activity changes. *Phys Med Biol*. 2000;45(12):3525–3543.
16. Imbert L, Poussier S, Franken PR, et al. Compared performance of high-sensitivity cameras dedicated to myocardial perfusion SPECT: a comprehensive analysis of phantom and human images. *J Nucl Med*. 2012;53(12):1897–1903.
17. Nesterov SV, Deshayes E, Sciagrà R, et al. Quantification of myocardial blood flow in absolute terms using 82Rb PET imaging: the RUBY-10 study. *JACC Cardiovasc Imaging*. 2014;7(11):1119–1127.
18. Lodge MA, Lucas JD, Marsden PK, Cronin BF, O'Doherty MJ, Smith MA. A PET study of 18FDG uptake in soft tissue masses. *Eur J Nucl Med*. 1999;26(1):22–30.
19. Wernick MN, Infusino EJ, Milosevic M. Fast spatio-temporal image reconstruction for dynamic PET. *IEEE Trans Med Imaging*. 1999;18(3):185–195.
20. Cui J, Liu X, Wang Y, Liu H. Deep reconstruction model for dynamic PET images. *PLoS One*. 2017;12(9):e0184667.
21. Hashimoto F, Ohba H, Ote K, Teramoto A, Tsukada H. Dynamic PET image denoising using deep convolutional neural networks without prior training datasets. *IEEE Access*. 2019;7:96594–96603.
22. Sanaat A, Shiri I, Arabi H, Mainta I, Nkoulou R, Zaidi H. Deep learning-assisted ultra-fast/low-dose whole-body PET/CT imaging. *Eur J Nucl Med Mol Imaging*. 2021;48(8):2405–2415.
23. Klyuzhin IS, Cheng JC, Bevington C, Sossi V. Use of a tracer-specific deep artificial neural net to denoise dynamic PET images. *IEEE Trans Med Imaging*. 2020;39(2):366–376.
24. Rubinstein E, Salhov M, Nidam-Leshem M, et al. Unsupervised tumor detection in Dynamic PET/CT imaging of the prostate. *Med Image Anal*. 2019;55:27–40.
25. Sanaat A, Arabi H, Mainta I, Garibotto V, Zaidi H. Projection-space implementation of deep learning-guided low-dose brain PET imaging improves performance over implementation in image-space. *J Nucl Med*. 2020;61(9):1388–1396.
26. Chen KT, Gong E, de Carvalho Macruz FB, et al. Ultra-low-dose (18)F-Florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology*. 2019;290(3):649-656.
27. Torizuka T, Nobezawa S, Momiki S, et al. Short dynamic FDG-PET imaging protocol for patients with lung cancer. *Eur J Nucl Med*. 2000;27(10):1538–1542.
28. Monden T, Kudomi N, Sasakawa Y, Yamamoto Y, Kawai N, Nishiyama Y. Shortening the duration of [18 F] FDG PET brain examination for diagnosis of brain glioma. *Mol Imaging Biol*. 2011;13(4):754–758.
29. Visser EP, Kienhorst LB, de Geus-Oei L-F, Oyen WJ. Shortened dynamic FDG-PET protocol to determine the glucose metabolic rate in non-small cell lung carcinoma. Paper presented at: IEEE Nuclear Science Symposium Conference Record. 2008; Dresden, Germany, Germany.
30. Samimi R, Kamali-Asl A, Geramifar P, van den Hoff J, Rahmim A. Short-duration dynamic FDG PET imaging: optimization and clinical application. *Phys Med*. 2020;80:193–200.
31. Lee AX, Zhang R, Ebert F, Abbeel P, Finn C, Levine S. Stochastic adversarial video prediction. arXiv preprint arXiv:180401523. 2018.
32. Babaeizadeh M, Finn C, Erhan D, Campbell RH, Levine S. Stochastic variational video prediction. arXiv preprint arXiv:171011252. 2017.
33. Denton E, Fergus R. Stochastic video generation with a learned prior. arXiv preprint arXiv:180207687. 2018.
34. Vondrick C, Torralba A. Generating the future with adversarial transformers. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). HI, USA: Honolulu; 2017.
35. Villegas R, Yang J, Hong S, Lin X, Lee H. Decomposing motion and content for natural video sequence prediction. arXiv preprint arXiv:170608033. 2017.
36. Lu C, Hirsch M, Scholkopf B. Flexible Spatio-temporal networks for video prediction. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA; 2017.
37. Vondrick C, Pirsiavash H, Torralba A. Generating videos with scene dynamics. Paper presented at: Advances In Neural Information Processing Systems; 2015.
38. Saito M, Matsumoto E, Saito S. Temporal generative adversarial nets with singular value clipping. Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017.
39. Chen B, Wang W, Wang J. Video imagination from a single image with transformation generation. arXiv preprint arXiv:170604124v2. 2017.
40. Tulyakov S, Liu M-Y, Yang X, Mocogan KJ. Decomposing motion and content for video generation. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
41. Larsen ABL, Sønderby SK, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv:151209300. 2015.

42. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:13126114. 2013.
43. Higgins I, Matthey L, Pal A, et al. beta-vae: Learning basic visual concepts with a constrained variational framework. International Conference on Learning Representations; 2017; Toulon, France.
44. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Paper presented at: Advances in neural information processing systems; 2014.
45. Zhu J-Y, Zhang R, Pathak D, et al. Toward multimodal image-to-image translation. Paper presented at: Advances in neural information processing systems; 2017.
46. Kingma DP, Adam BJ. A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.
47. García-Gómez FJ, García-Solís D, Luis-Simón FJ, et al. Elaboration of the SPM template for the standardization of SPECT images with 123I-Ioflupane. *Rev Esp Med Nucl Imagen Mol*. 2013;32(6):350–356.
48. Patlak C, Dhawan V, Takikawa S, Chaly T. Estimation of striatal uptake rate constant of FDOPA using PET: methodological issues. Vol 1030: Elsevier Science Publishers;. 1993.
49. Whone AL, Bailey DL, Remy P, Pavese N, Brooks DJ. A technique for standardized central analysis of 6–18F-fluoro-L-DOPA PET data from a multicenter study. *J Nucl Med*. 2004;45(7):1135–1145.
50. Schoonjans F, Zalata A, Depuydt C, Comhaire F. MedCalc: a new computer program for medical statistics. *Comput Methods Programs Biomed*. 1995;48(3):257–262.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Sanaat A, Mirsadeghi E, Razeghi B, Ginovart N, Zaidi H. Fast dynamic brain PET imaging using stochastic variational prediction for recurrent frame generation. *Medical Physics*. 2021;48:5059–5071. https://doi.org/10.1002/mp.15063