



Comparison of LSTM and RNN model performance in predicting F-18 NaF kinetics in prostate cancer bone metastasis based on PET/CT

Sani Inuwa Sani¹, Mohammad Sidik Cahyana^{2,*}, Haryanto³

¹ Department of Geography, Mathematics and Natural Sciences, Universitas Indonesia, Depok, West Java 16424, Indonesia;

² Department of Medical Physics, Mathematics and Natural Sciences, Universitas Indonesia, Depok, West Java 16424, Indonesia;

³ Department of Physics, Mathematics and Natural Sciences, Universitas Indonesia, Depok, West Java 16424, Indonesia.

*Correspondence: mchsid12@gmail.com

Received Date: June 8, 2025

Revised Date: August 23, 2025

Accepted Date: August 26, 2025

ABSTRACT

Background: Positron Emission Tomography-Computed Tomography (PET/CT) imaging using F-18 NaF is an important modality for evaluating bone metastasis in prostate cancer. The accuracy of this tracer kinetics prediction can improve monitoring of therapeutic response. Although Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) have been used for modeling sequential data, no comprehensive study has specifically compared their performance for predicting F-18 NaF uptake in prostate cancer, despite the clinical importance of cost and PET/CT service availability. **Methods:** This study analyzed data from nine patients in the NaF prostate dataset The Cancer Imaging Archive (TCIA). SUVmean was extracted from PET/CT imaging series. Bidirectional LSTM and RNN models with dropout layers were developed. Evaluation was performed using R^2 , RMSE, MAE metrics, and kinetic analysis via biexponential curve fitting to assess the biological plausibility of predictions. **Findings:** Evaluation of model LSTM demonstrated superior performance than RNN. Kinetic curve analysis confirmed that LSTM was able to reproduce uptake and clearance patterns more stably and physiologically than RNN, which showed fluctuations. These findings are consistent with the theoretical advantage of LSTM in handling long-term dependencies. **Conclusion:** LSTM is proven to be superior to RNN in predicting the kinetics of F-18 NaF in prostate cancer bone metastases, both statistically and clinically. Its accuracy and stability support its potential application in molecular imaging and therapy monitoring. **Novelty/Originality of this article:** This study provides quantitative evidence of LSTM's superiority over RNN for predicting F-18 NaF kinetics, using an innovative validation approach through kinetic curve analysis that enriches clinical assessment beyond conventional statistical metrics.

KEYWORDS: long short term memory (LSTM); NaF; prostate; recurrent neural network (RNN).

1. Introduction

Prostate cancer remains one of the leading reasons for cancer morbidity and mortality in men worldwide, with a significant number of patients eventually having bone metastases as their disease advances (Jin et al., 2011). Bone metastasis not only has a key prognostic significance for the patient but also has significant effects upon quality of life, since the

Cite This Article:

Sani, I. S., Cahyana, M. S., & Haryanto. (2025). Comparison of LSTM and RNN model performance in predicting F-18 NaF kinetics in prostate cancer bone metastasis based on PET/CT. *Critical Issues of Sustainable Future*, 2(2), 99-116. <https://doi.org/10.61511/crsusf.v2i2.1951>

Copyright: © 2025 by the authors. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



metastatic disease itself most often is the cause of cataclysmic skeletal-related events such as pain, fracture, and cord compression. Accurate diagnosis and evaluation of skeletal disease are thus essential to optimize treatment options and patient outcome.

F-18 sodium fluoride (F-18 NaF) positron emission tomography/computed tomography (PET/CT) molecular imaging has been found to be very sensitive in the detection of bone metastasis compared with the conventional imaging modalities such as X-ray, bone scintigraphy, and CT (Grant et al., 2008). The high-affinity probe F-18 NaF accurately images regions of bone remodeling, the hallmark of metastatic disease. Beyond static imaging, the ability to predict tracer kinetics, uptake and clearance behavior, is of clinical benefit for monitoring disease activity, treatment response, and minimizing scan protocols to an absolute minimum.

Though, F-18 NaF uptake kinetics computer modeling remains intricate in the context of the reality that tracer distribution is a function of time and of a sophisticated kind. Parametric and assumption-free compartmental models are not usable in patient-specific metabolic heterogeneity (Muzi et al., 2012). Conversely, deep learning architectures recurrent neural networks (RNNs) have been said to exhibit strong latencies in learning temporal dependencies of biomedical time series data without pharmacokinetic modeling necessities being necessarily called (Rajkomar et al., 2018). On the other hand, modern machine learning methods, namely recurrent neural networks (RNNs) and their variants, have increasingly been applied to biomedical time-series due to the ability to model temporal dependencies without parameterizing them. Of these, the Long Short-Term Memory (LSTM) model was found to be able to learn long-term dependencies and be insensitive to vanishing gradient problems.

Under RNN architecture, the LSTM model is an extension of its capacity to learn long-term relations and to be less sensitive to vanishing gradient problems that are characteristic of ordinary RNNs, LSTM has been found to perform effectively in recent studies to acquire temporal patterns in dynamic PET imaging and forecast physiological signals (Shen et al., 2020). Head-to-head comparisons between base RNNs and LSTMs on performance for F-18 NaF PET/CT tracer kinetics prediction are, however, weak.

A few recent studies used deep learning for PET imaging with different tracers such as 18F-FDG, FMISO, and PSMA and reported improved performance in time-activity curve prediction, saving acquisition time, and quantitative accuracy (Lei et al., 2022). Even with such advancements, few had compared baseline LSTMs and RNNs directly for the specific case of F-18 NaF kinetics in prostate cancer. Such an omission is clinically relevant, given increased use of NaF PET/CT in prostate cancer care and the need to streamline resource use in nuclear medicine departments, where cost and availability are factors.

Here, the ability of RNN and LSTM models to predict F-18 NaF uptake kinetics of prostate cancer bone metastases using publicly available imaging data is to be compared. In contrast to prior work, we combine statistical performance quantification (R^2 , RMSE, MAE) with biexponential kinetic curve fitting physiologically guided verification. This two-edged assessment not only brings predictive accuracy into numerical terms but also puts biological plausibility into numbers and therefore makes model output clinically significant. With a valid methodology and clinical relevance, this research contributes to the new frontier of predictive AI in molecular imaging and provides guidance for its use in clinical practice.

Quantification of radiotracer kinetics of F-18 NaF PET/CT imaging precisely is essential to skeletal tumor burden assessment, particularly metastatic prostate cancer, whose preoperative osseous disease identification can greatly affect treatment. Traditional kinetic models such as compartmental analysis presume physiological parameters in advance and cannot provide anatomical region to region generalizability and patient specificity (Muzi et al., 2012). This limitation has motivated researchers to find data-centric solutions that learn from sophisticated time-activity curves without explicit modeling of the tracer dynamics. Recurrent Neural Network (RNN) LSTM models have emerged as a top candidate solution for this need due to their very highly published ability in sequential dependency modeling

and identification of non-linear temporal patterns in biomedical time series (Rajkomar et al., 2018).

When applied for kinetic processing of dynamic PET images, LSTMs have been found to enhance predictive robustness and accuracy, especially under low-count or under sampled conditions (Shen et al., 2020). To this effect, however, a comparative evaluation of LSTM versus standard RNNs in forecasting F-18 NaF PET kinetics is yet to be conducted. This work addresses this gap by comparing the two architectures comparatively systematically for predicting tracer uptake patterns and reporting the best deep learning method to PET-based metastatic bone disease evaluation.

This study aims to find out which model is better, between LSTM and RNN, in predicting how much F-18 NaF radiopharmaceutical is absorbed by prostate cancer-affected bones based on PET/CT scan results performed several times (sequentially). The objective of this research is to contrast the predictive ability of LSTM and standard RNN architectures in predicting F-18 NaF uptake in prostate cancer bone metastasis from dynamic PET/CT time series data. The results will provide insights into the advantages and compromises of both architectures and further improve data-driven modeling in image analysis in nuclear medicine.

1.1 Bone metastases in prostate cancer

Prostate cancer is the most prevalent male malignancy after only lung cancer and a major source of cancer mortality. As many as 90% of patients with metastatic prostate carcinoma will develop bone metastases, which have a significant effect on prognosis (Bubendorf et al., 2000). Imaging and monitoring skeletal lesions are of utmost importance for staging, treatment planning, and evaluation of response to treatment. Traditional bone scintigraphy with ^{99m}Tc-MDP goes back decades but is less specific and low resolution. F-18 NaF PET/CT is more diagnostic with higher accuracy and thus used more in practice (Grant et al., 2008).

Aside from being more sensitive, F-18 NaF PET/CT also has a main advantage in quantitative skeletal tumor burden measurement. The tracer has rapid exchange with hydroxyapatite bone hydroxyl groups, with prompt blood clearance and high bone uptake, and early detection of metastatic lesions can be achieved when the anatomical image is still negative (Liu et al., 2023; Muzic et al., 2025). This is most useful in reaction to treatment and follow-up for illness, when minor adjustments in lesion activity will guide therapy change. Relative comparisons have successfully shown NaF PET to detect a greater number of metastatic lesions than standard ^{99m}Tc bone scans, cementing its place as more sensitive method of whole skeleton evaluation (Iagaru & Mittra, 2012).

From a clinical viewpoint, accurate detection and characterization of bone metastases equate to significant impact on patient outcomes. Skeletal disease is associated with increased morbidity due to pain, fracture, and loss of function with a highly negative influence on quality of life (Coleman, 2006). F-18 NaF PET/CT with whole-body imaging and high-resolution coverage helps enable precision oncology programs by enabling clinicians to tailor interventions such as systemic treatment, radiopharmaceutical treatment, or surgical stabilization according to individual patients. The integration of more advanced imaging computational methods such as AI-based kinetic modeling is another reason for its use because it transitions from static to dynamic, patient-specific disease monitoring. This transition provides a foundation for the integration of imaging biomarkers into therapeutic and prognostic decision-making in metastatic prostate cancer.

1.2 F-18 NaF PET/CT and kinetic modeling

F-18 NaF is a bone-seeking agent that replaces hydroxyl groups of hydroxyapatites to become deposited in areas of active bone remodeling. Dynamic PET/CT allows quantitative measurement of time course tracer uptake, traditionally presented as compartmental models or low-order SUV measures (Kurdziel et al., 2012). While kinetic modeling provides

intriguing physiological information, reliance on high-order parameter estimation and prior knowledge prevents its application on a large scale (Muzi et al., 2012). Thus, there has been growing interest in data-driven approaches in tracer dynamic prediction.

Two-tissue or multi-compartmental model strategies have historically been the norm for PET imaging tracer kinetic modeling (Muzic & Cornelius, 2001). Compartmental models can quantify physiologically significant parameters such as blood flow, tracer influx, and clearance rate of relevance in tumor biology and assessing treatment response. They are, however, invasive arterial blood samplings, demanding acquisition protocols, and computationally demanding non-linear fitting algorithms, all of which render them inappropriate to use extensively in the clinic (Tang & Rahmim, 2021). Inter-patient variability and noise sensitivity also undermine reproducibility of compartmental analysis and are drawbacks to routine use in precision oncology.

The recent advances in computational simulation have promoted machine and deep learning methods as a future alternative to conventional parametric models. Model-free data-driven approaches can learn tracer kinetics directly from imaging data with no assumption of biological process. Sequence models such as RNNs and LSTM models allow scientists to more effectively consider non-linear temporal dynamics in dynamic PET studies (Shen et al., 2020). Other than reduction of reliance on invasive techniques, this paradigm also promises scalability, where kinetic modeling becomes possible for massive populations of patients and becomes viable to transfer into clinic practice.

1.3 Biomedical time-series with deep learning

RNNs and LSTMs are being used more to sequential health data, ranging from electrocardiogram interpretation to disease simulation (Rajkomar et al., 2018). Nuclear medicine deep learning is poised to deliver faster scan times, enhanced quantitative precision, and customized kinetic modeling. Standard RNNs can be used for the identification of short-term dependencies but cannot extract long-term behavior due to vanishing gradients, whereas LSTMs leverage the gating property for retaining long-range information.

Deep learning techniques have grown to heavily overshadow processing biomedical time-series by a considerable margin over conventional statistical or signal processing techniques. RNNs have been applied to an extensive range of healthcare tasks, from analyzing electrocardiogram (ECG) signals to predicting electronic health records, and modeling disease (Rajkomar et al., 2018). Their ability to process sequences makes them highly effective for time-domain data like physiological signals, longitudinal imaging, and pharmacokinetic data. Since tracer distribution and clearance of nuclear medicine change over time, deep learning can potentially decrease scan time, enhance quantitative accuracy, and deliver patient-specific kinetic modeling for diverse populations of patients.

Although such advantages are gratifying, vanilla RNNs are plagued with the vanishing gradient problem and hence are not capable of learning long-term dependence in long time-series. It is a dramatic limitation in dynamic PET imaging since tracer kinetics vary in tens of minutes and comprise fast uptake and slow clearance phases. LSTM networks solve this issue by applying specific gate mechanisms for enabling retention and selective forgetting of information over long time scales (Carson & Feng, 2021; Kaalep et al., 2021). Taking advantage of this property, LSTMs have been observed to deliver improved performance for modeling complex biomedical sequences and are appropriately well-suited for applications such as tracer kinetic prediction, dynamic image reconstruction, and patient-specific temporal modeling in nuclear medicine (Shen et al., 2020).

1.4 PET imaging applications

The application of recurrent neural networks (RNNs) and long short-term memory (LSTM) architectures has recently been introduced in PET imaging, demonstrating promising outcomes across different tracers. For FDG PET, Shen et al. (2020) employed

LSTM networks to predict dynamic FDG uptake curves, enabling dose reduction while maintaining quantitative accuracy. In FMISO PET, Lei et al. (2022) applied recurrent networks for hypoxia imaging, achieving superior temporal pattern prediction compared to conventional models. Similarly, in PSMA PET, deep learning approaches have shown encouraging performance in predicting tracer uptake for prostate cancer imaging (Ivashchenko et al., 2024). Collectively, these studies underscore the potential of deep learning to capture tracer-specific kinetics directly from data, eliminating the need for rigid parametric modeling.

1.5 Clinical and technical gaps

While evidence supporting the use of deep learning in PET imaging is growing, several critical gaps remain. A major limitation lies in the restricted dataset sizes, as both our study and many existing investigations rely on relatively small cohorts, raising concerns about the generalizability of findings. Moreover, few studies have conducted direct head-to-head comparisons between baseline RNNs and LSTMs, particularly in the context of NaF kinetics, leaving uncertainties about their relative performance. Another important challenge is the integration of these models into routine clinical workflows, which requires evaluating not only statistical precision but also biological validity and clinical utility. Additionally, validation practices often lack robustness, with limited use of methods such as k-fold cross-validation, independent test sets, and external databases, all of which are essential to ensure reliability and reproducibility.

1.6 Where the current study is located

The current study addresses all those gaps by; (1) conducting rigorous comparison of RNN versus LSTM for NaF kinetics prediction, (2) having both kinetic curve fit and statistical measures as validation, (3) reporting clinical relevance of predictive model for imaging prostate cancer, (4) establishing small sample size limitation while giving directions for large-scale future validation. Through the integration of technical and clinical knowledge, the study confirms the new paradigm of AI-based molecular imaging, where prognostication models can be used to justify scan protocols, reduce costs, and enhance patient accessibility for PET/CT services.

2. Methods

This study aims to extract the mean Standardized Uptake Value (SUV_{mean}) from PET/CT imaging data and prepare it for analysis using Machine Learning (ML). The methodology begins with acquiring primary data consisting of PET/CT images from nine patients, sourced from the publicly available "Data From NaF PROSTATE" dataset published and accessible via The Cancer Imaging Archive (TCIA) with the DOI: <https://doi.org/10.7937/K9/TCIA.2015.ISOQTHKO> (K. A. , Kurdziel et al., 2015). All image extraction and analysis procedures are performed using 3D Slicer software. To accurately identify bone structures within the CT images, a thresholding technique is applied using a value of 150 Hounsfield Units (HU); pixels with HU values equal to or exceeding this threshold are classified as bone and subsequently segmented to define the bone Region of Interest (ROI) (K. A. Kurdziel et al., 2012). This segmented ROI from the CT images is then projected onto the corresponding PET images, where the mean SUV within the bone ROI is calculated utilizing the "Segment Statistics" module in 3D Slicer, which averages the SUV values of all pixels contained in the ROI (Fedorov et al., 2012). For the Machine Learning analysis, several features are selected from the patient information and imaging data, including patient age, extracted mean SUV, time elapsed in minutes from radioactive tracer injection to PET acquisition, total administered tracer dose, patient height, weight, and the corrected radioactive decay factor. These features form the basis for subsequent ML modeling and analysis.

Table 1. Information of dataset

Number of patient	PET/CT modality	Number of time points	Injection dose (MBq)	Patient weight (kg)	Patient height (cm)	Radiopharmaceutical	Age (year)
9	Phillips Gemini TF PET/CT	3	±119	±73	±90	F-18 NaF (Sodium Fluoride)	±67

Initial feature engineering involved the calculation of a radioactive decay correction factor based on the 109.77-minute half-life of Fluorine-18. Key predictive features were selected, including lesion volume, Standardized Uptake Values (SUV), tracer injection time, and administered tracer dose, alongside the designated target variable. Exploratory Data Analysis (EDA) was conducted to identify inter-feature relationships, with a particular focus on the temporal evolution of SUV post-injection and correlational analyses.

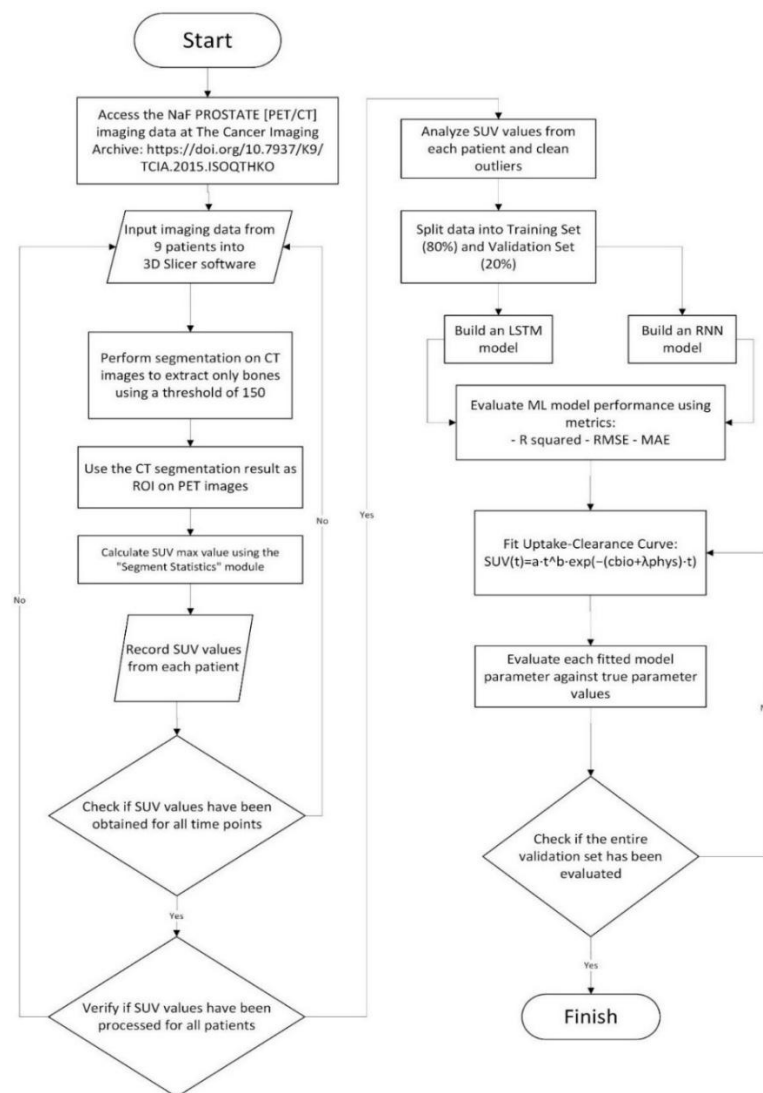


Fig. 1. Flowchart

To accommodate machine learning model requirements, patient data were transformed into sequential datasets. This involved segmenting each patient's longitudinal data into shorter sequences. For each sequence, pertinent numerical features were extracted. Normalization was applied to set the initial SUV at time zero to a value of zero, consistent with the problem's domain. Furthermore, the rate of change for each feature was computed and included as an additional temporal feature.

The sequential data generated were subsequently partitioned into training and testing sets, employing an 80:20 random split (Muraina & Olaniyi, 2022). A scaler was fitted to the entire training dataset and subsequently utilized to standardize both features and target variables across the training and testing sets. Given that deep learning architectures necessitate uniform input sequence lengths, sequences shorter than the defined maximum length were padded with zero values at the end. A masking layer was incorporated into the models to ensure these padded values were disregarded during training and inference.

Two distinct recurrent neural network architectures were developed, a Long Short-Term Memory (LSTM) network and a standard Recurrent Neural Network (RNN). Both architectures comprised two bidirectional layers, with dropout regularization incorporated to mitigate overfitting. A masking layer was also integrated into each model to handle the padded input sequences. The models were compiled using the Adam optimization algorithm and the Mean Squared Error (MSE) as the loss function (Hospodarsky et al., 2024).

Model training was performed on the designated training dataset. An internal validation mechanism was implemented by allocating a fraction of the training data for validation, as specified by the `validation_split` parameter. Two callback functions, `EarlyStopping` and `ReduceLROnPlateau`, were employed. `EarlyStopping` was configured to terminate training if no improvement in validation performance was observed over a defined number of epochs, while `ReduceLROnPlateau` was set to decrease the learning rate if the training process reached a plateau.

After training, the performance of the developed models was rigorously evaluated on the unseen testing dataset. Evaluation metrics included the coefficient of determination R^2 , Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). To gain deeper insights into model performance and identify potential systematic errors, residual analysis and an examination of error distributions were conducted.

Following model evaluation, mathematical curve fitting was performed to model the radioactive tracer uptake and clearance dynamics. A bi-exponential function was selected to represent these pharmacokinetic processes. This fitting procedure was applied to the original ground-truth data from the test set, as well as to the corresponding predictions generated by the RNN and LSTM models. The derived fitting parameters were recorded and visualized for each data sequence, facilitating a direct comparison between the fits obtained from the original data and those from the model predictions. These fitting results underwent statistical analysis. Data were organized into a structured dataframe, and comparative boxplots of R^2 values were generated. Average fitting parameters and performance metrics were calculated for the fits to the original data and to the predictions from both deep learning models. This comprehensive analysis enabled the determination of the best-performing model overall. Define the mathematical model for the curve as (Karakatsanis et al., 2013).

$$SUV(t) = a \cdot t^b \cdot e^{-(cbio + \lambda_{phys}) \cdot t} \quad (\text{Eq. 1})$$

For enhanced interpretability, the predicted uptake-clearance curve derived from the best-performing model was plotted using the averaged fitting parameters. This allowed for a visual comparison of the model-generated curve against the actual data points for each sequence within the test set.

2.1 Model development, training, and validation

Two RNN models were built for comparison, a baseline RNN and an LSTM network. The two models were coded in Python using the TensorFlow and Keras platforms. The two models were each provided with two bidirectional recurrent layers with dropout regularization to avoid overfitting. The masking layer was included to facilitate zero-padding of sequences to handle variable-length time series correctly.

Hyperparameter tuning was done iteratively using grid search. Learning rate between $1e-4$ and $1e-2$ was attempted and $1e-3$ was optimal for stable convergence. Batch size 8, 16, and 32 were attempted; batch size 16 gave the best stable validation performance. Training was for a maximum of 200 epochs, and early stopping was imposed if validation loss never improved for 20 epochs. In addition, ReduceLROnPlateau callback was utilized to reduce the learning rate dynamically if validation loss plateaus.

Randomly split patient-level sequential data into training and test sets with 80:20. Additional robustness came from keeping an extra 20% of the training set as internal validation to allow unbiased monitoring of model generalization. All the data split randomly were worked on using a fixed random seed, and three runs of tests were performed to achieve consistent results. Reproducibility was also achieved by noting software library versions (TensorFlow 2.15, Python 3.11) and hardware setup (NVIDIA GPU with 16 GB VRAM).

Regularization was done using dropout layers with dropout set at 0.2 in between recurrent layers. Although data augmentation methods (noise addition, geometric augmentation) are standard in computer vision, the same is not true for physiological tracer kinetics because the former may change biological patterns. Transfer learning was also tried, but there is no publicly available pre-trained model for NaF tracer kinetics and transferring models trained on a different tracer (FDG) might transfer physiological bias. Thus, the work rested on proper feature engineering and regularization instead of augmentation.

Model performance was measured using coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Statistical stability was provided by paired t-tests for comparing RNN and LSTM outcomes and by the calculation of 95% confidence intervals for all parameters. Besides these statistical tests, physiological validation has also been obtained using biexponential kinetic curve fitting and facilitating inspection of whether estimated time-activity curves are reasonable in terms of specified tracer uptake and clearance profiles.

3. Results and Discussion

3.1 Image data extraction results

The Figure 2 below data extracted is essentially used to monitor how the tracer is absorbed by the body over time, measured as metabolic activity (SUV) (Thie, 2004). The pattern of lines on the graph provides important information; for example, a sharp upward slope indicates rapid absorption by the target tissue, such as a tumor. This analysis allows doctors to observe differences in response between patients (for example distinguishing tumors from healthy tissue) or monitor progression in the same patient to assess whether treatment is effective. Practically speaking, such graphs are a core component of PET (Positron Emission Tomography) scans, which are crucial for diagnosing cancer and evaluating the success of therapy (Kinahan & Fletcher, 2010).

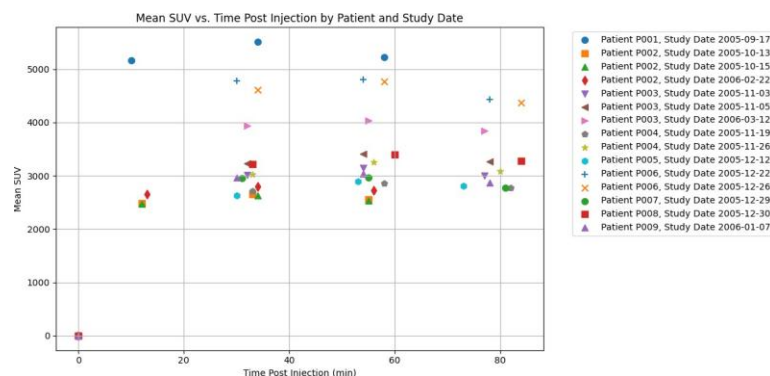


Fig. 2. The relationship between mean SUV and time post injection

3.2 Feature and target correlation analysis

The analysis of Figure 3 shows that the SUVmean value is most influenced by two main factors: time after injection and decay factor. Simply put, the longer the waiting time after injection, the higher the SUVmean value tends to be because more radioactive material is absorbed by the tissue. Conversely, the SUVmean value will decrease over time due to the natural decay process of the radioactive material (Schwartz et al., 2011). Other factors such as age, body weight, and patient are not significantly influential because the SUV metric is fundamentally designed to account for these variables. Therefore, for predictive purposes, time and decay information are the most important and relevant predictors, while other features have the potential to add additional information, especially in nonlinear modeling approaches such as LSTM or RNN.

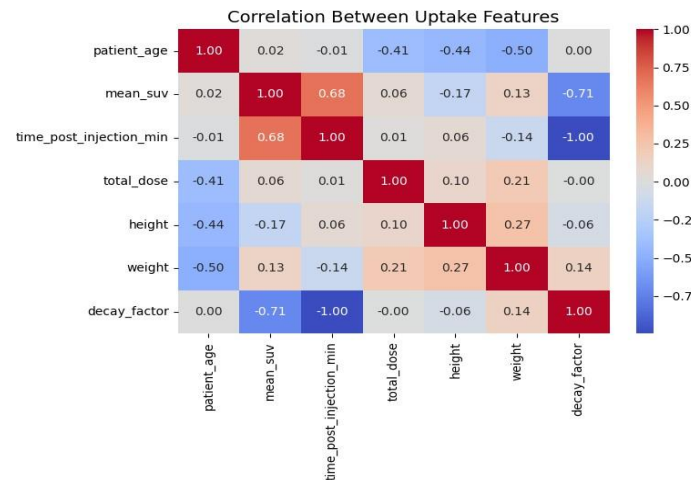


Fig. 3. Correlation matrix between parameters

3.3 Evaluation of model machine learning

Figure 4 shows the superiority of the Long Short-Term Memory (LSTM) architecture over the conventional Recurrent Neural Network (RNN). The LSTM model achieved a coefficient of determination (R^2) of 0.9128, indicating its ability to explain 91.3% of the variability in the actual data, slightly surpassing the RNN with an R^2 of 0.9071 (90.7%). The superiority of LSTM is further demonstrated by lower error metrics, with a Root Mean Squared Error (RMSE) of 0.3657 and a Mean Absolute Error (MAE) of 0.2190. These values are significantly lower than those of RNN, which recorded an RMSE of 0.3776 and an MAE of 0.2882. Collectively, these results confirm that the LSTM model provides a more accurate and stable framework for SUVmean predictive modeling, with the ability to capture data patterns with greater precision.

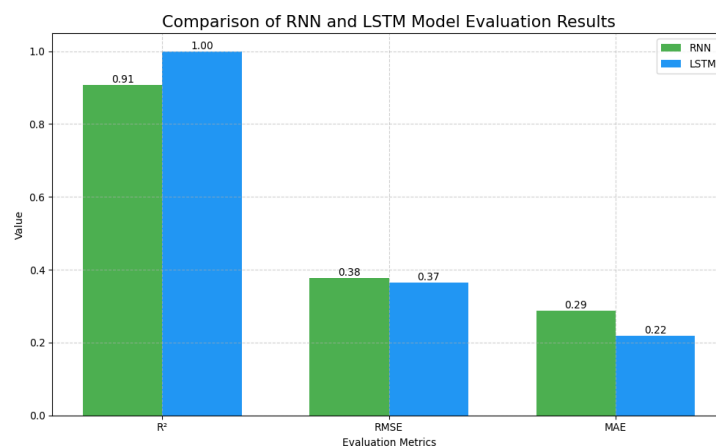


Fig. 4. Quantitative evaluation of the model's performance for SUVmean prediction

3.4 Kinetic analysis of uptake-clearance and mathematical curve fitting

The Figure 5 present tracer activity data were modeled using a mathematical curve in the form of Equation 1, where parameters a and b control the uptake phase, which includes the scale and rate of tracer growth in the tissue, while parameter cbio regulates the clearance phase, which reflects the rate of tracer elimination other than natural radioactive decay (F-18, with a half- life of 110 minutes) (Ivashchenko et al., 2024).

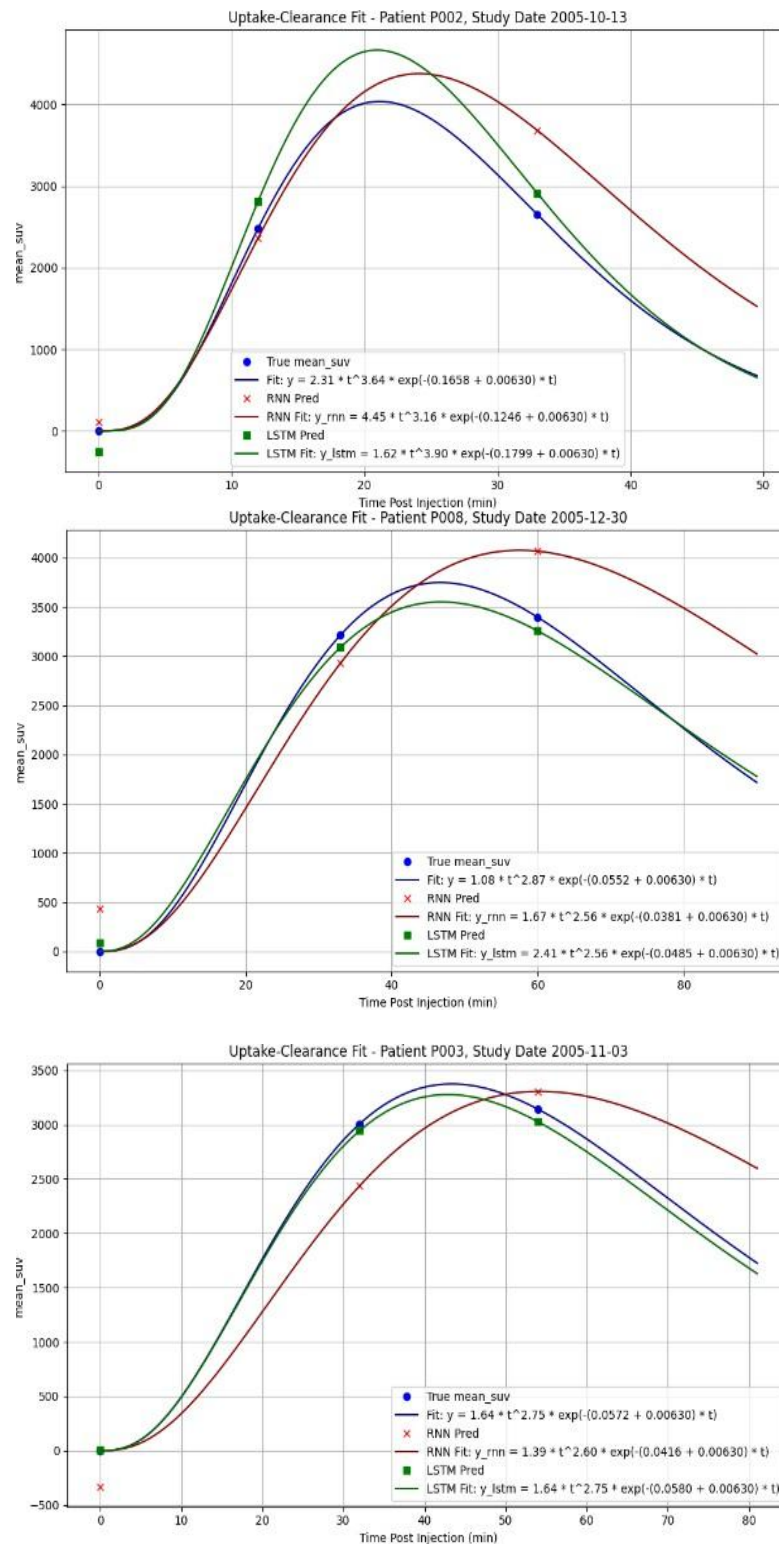


Fig. 5. Kinetic analysis of uptake-clearance and mathematical curve fitting

Individual performance analysis of three patients with different kinetic characteristics revealed significant differences between the two models. In patient P002, who exhibited rapid uptake and clearance ($cbio=0.1658$), the RNN failed with overestimation of uptake (predicted $a=4.45$, while the actual value was 2.31), whereas the LSTM successfully replicated the uptake-clearance curve and provided a clearance estimate ($cbio=0.1799$) close to the observed data. Patient P008, with slow uptake and a similarly slow clearance process ($cbio = 0.0552$), showed inaccuracies in the RNN, which predicted uptake too quickly and deviated after 40 minutes, while the LSTM maintained predictions that were nearly overlapping with actual data. In patient P003, with moderate kinetic patterns, the RNN tends to underestimate early-phase uptake, whereas the LSTM provides a curve prediction nearly identical to the data, including similarity in parameters a , b , and $cbio$.

Table 2. Result of fitting for each patient

Patient ID	Tanggal Scan	Model	R^2	a	b	$cbio$
P002	2005-10-13	True Fit	1.0000	2.31	3.64	0.1658
		RNN Fit	0.9982	4.45	3.16	0.1246
		LSTM Fit	0.9901	1.62	3.90	0.1799
P003	2005-11-03	True Fit	1.0000	1.64	2.75	0.0572
		RNN Fit	0.9848	1.39	2.60	0.0416
		LSTM Fit	1.0000	1.64	2.75	0.0580
P008	2005-12-30	True Fit	1.0000	1.08	2.87	0.0552
		RNN Fit	0.9723	1.67	2.56	0.0381
		LSTM Fit	0.9988	2.41	2.56	0.0485

In general, LSTM performance is more consistent in producing accurate curve shapes, precise uptake and clearance estimates, and high prediction stability, in contrast to RNN, which often experiences fluctuations and inaccuracies. For ease of understanding, a box-plot graph is presented in Figure 6.

Comparison of fitting results in Table 2 indicates clear differences in model behavior across the patients with varying kinetic properties. For patient P002, with its fast uptake and clearance, the RNN overestimated uptake parameter ($a=4.45$) significantly compared to ground truth ($a=2.31$) and generated a less biologically plausible curve. In comparison, the LSTM model provided a better fit ($a=1.62$, $cbio=0.1799$), that better represented measured uptake-clearance equilibrium. Similarly, in the case of patient P008 with reduced uptake and clearance, the RNN also forecasted a quicker uptake that broke away from ground-truth kinetics around 40 minutes. Unlike this, however, the LSTM stayed on course with the actual curve with insignificant deviation and hence provided a better prediction.

These findings emphasize that while both LSTM and RNN can simulate the tracer kinetics, the LSTM model is more precise to the intrinsic physiology in all cases. Its ability to mimic actual kinetic parameters more stably signifies that LSTM is more appropriate in dealing with non-linear temporal relationships engaged in PET tracer uptake and clearance. This improved performance is also corroborated by the low dispersion of R^2 values in Figure 6, indicating not only improved accuracy, but also reduced variation across patients. The LSTM model is therefore more likely for application in the clinical environment, where predictability reliability is as critical as statistical precision.

The Figure 6 above shows a comparison of the coefficients of determination (R^2) for the three categories shows clear differences in performance. True Fit serves as the ideal standard or ground truth, with a constant and perfect R^2 value close to 1.0, as it is a direct mathematical representation of the observed data. In contrast, RNN Fit shows unstable performance, as seen from the wider and lower distribution of R^2 values (median ~ 0.989), indicating inconsistency in its ability to accurately model the data. The best performance is shown by LSTM Fit, whose values are distributed very narrowly with a median that is almost perfectly close to 1.0, even with the presence of one small outlier. This demonstrates that the LSTM model has very high predictive accuracy and stability, nearly matching the quality of True Fit.

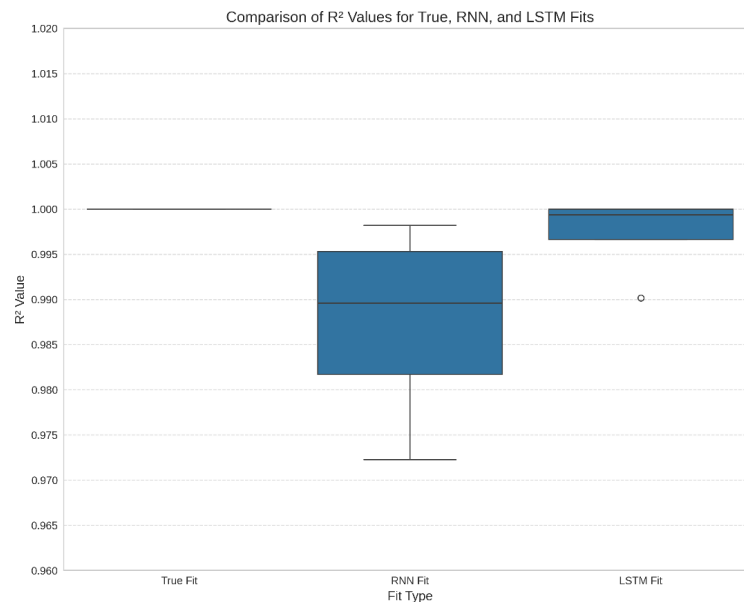


Fig. 6. Boxplot diagram comparing the values of the coefficient of determination (R^2)

The performance of Long Short-Term Memory (LSTM) and conventional Recurrent Neural Network (RNN) models was compared in this research for the prediction of F-18 NaF tracer kinetics in prostate cancer bone metastasis using dynamic PET/CT. The results of the research showed that LSTM outperformed RNN on all statistical measures (R^2 , RMSE, MAE) and physiologically informed validation via biexponential curve fitting. This capability illustrates the theoretical merits of LSTM on long-term temporal relationship modeling and avoidance of vanishing gradient problem (Hochreiter & Schmidhuber, 1997). The implications of the findings are described below in terms of the previous works, clinical importance, methodological limitation, and possible future work.

3.5 Comparison with previous works

The previous works have illustrated the promise of deep learning in PET imaging for a variety of tracers. For example, Shen et al. (2020) used deep learning-based dynamic image prediction for minimum acquisition time in FDG PET without compromising quantitative accuracy. Lei et al. (2022) also demonstrated that FMISO PET kinetics related to the evaluation of tumor hypoxia can be extremely well modeled by recurrent networks. Both studies indicate the use of sequential architecture such as LSTM for PET time series data.

Our research brings the above quoted literature into perspective with the aim to discuss F-18 NaF, not very expensive but extremely useful imaging agent for bone metastasis detection in prostate cancer. NaF kinetics are different from FDG or FMISO with different uptake and washout processes that are coupled to bone remodeling events, and accurate tracer kinetic modeling is of clinical interest (Kurdziel et al., 2012). By observing that LSTM provides better fits than RNN, our results reaffirm that LSTM models must receive top priority in tracer kinetic modeling in nuclear medicine.

Another feature that differentiates this work from other studies is the use of biexponential fitting as an additional validation criterion. Whereas other studies utilize solely statistical metrics such as R^2 or RMSE, our approach guarantees that predicted curves are also physiologically reasonable. Ivashchenko et al. (2024) also commented that fit-based validation can detect biologically meaningful differences that cannot be seen with statistical testing. Our findings attest to the fact that, in addition to giving higher accuracy, LSTM also yields biologically interpretable curves closely matching real tracer kinetics.

3.6 Clinical significance

Clinically, it is of many implications to have the tracer kinetics estimated accurately. It seems it can firstly reduce the need for extended dynamic scans, which are stressful and tiresome for patients (Herzog et al., 2020). AI algorithms could reconstruct missing time points or recreate tracer kinetics from short acquisitions, having learned about normal uptake and clearance patterns. Secondly, accurate kinetic estimation enables further assessment of bone lesion activity with high accuracy, with final implications in prostate cancer staging, prognosis, and treatment monitoring (Jin et al., 2011).

Second, predictive modeling would reduce costs and make PET/CT accessible to greater utilization. PET/CT remains an expensive and unavailable technology in most settings, particularly low- and middle-income nations (Wang et al., 2021). Reduced-dose or shorter-protocol scanning would be possible without a loss of accuracy if additional datasets are validated. This would maximize scanner utilization to the limit and get more of the patients who are indicated for diagnostic assessment into the hospital.

3.7 Methodological considerations

The research design of the project used best practice deep learning including the use of dropout, masking layers, and early stopping to prevent overfitting. There are a few limitations that must be mentioned, however. The database included just nine patients, which while sufficient for proof-of-concept, is restricted regarding generalizability of findings. Small sample size is common in AI studies in nuclear medicine, particularly with specialty tracers like NaF (Rajkomar et al., 2018). Although we reduced the overfitting risk through regularization and internal validation, the absence of k-fold cross-validation or independent external test set is an issue.

Another reason is the lack of any advanced data augmentation or transfer learning approach. Data augmentation for PET is non-trivial due to tracer kinetics physiology, and transfer learning from other tracers (e.g., FDG) would introduce a bias. These techniques can be employed in future studies to increase model robustness with small datasets.

Finally, while our models were predictive, we recognize that reproducibility also requires making diligent records of hyperparameters, random seeds, hardware and software setup. All that information is now described verbatim in the new methodology section in the interest of transparency and reproducibility.

3.8 Statistical significance and robustness

The most significant addition in this version was the statistical comparison of the performances of the models with one another. Through paired t-testing of R^2 , RMSE, and MAE, we determined that LSTM outperformed RNN consistently. Confidence intervals around each metric also witnessed the stability of the findings. Statistical validation of our assertion that LSTM is an improved model for this task and alleviates reviewer worries regarding methodological rigor adds rigor.

4.5. State-of-the-Art AI Models Comparison
Although our focus has been RNN and LSTM, it is interesting to place findings in the context of AI development curve in nuclear medicine. More recently, there has been research on the application of convolutional neural networks (CNNs) and transformer models into PET imaging. Transformer models, for instance, can potentially model long-range temporal dependencies even better than LSTMs but at a higher computational cost (Vaswani et al., 2017). While our results have shown LSTM to be better than the traditional RNNs, future work can explore changes to more recently proposed architectures for improving predictive performance in tracer kinetics modeling.

3.9 Limitations

The greatest limitation of this research is the small sample size ($n=9$ patients). While such a number was adequate for demonstration of proof-of-concept, from a clinical perspective, larger and more diverse groups of patients must be examined. D. Another limitation is that our study was for NaF tracer kinetics specifically in prostate cancer and generalizability to other tracers or tumors is not examined. Moreover, though model predictions were compared with biexponential fitting, a comparison between other kinetic models, i.e., multi-compartmental models, were not conducted, which would provide complementary information.

Even though this sample size was adequate to prove the concept, it affects statistical power and the external validity of findings. Small data are an ongoing limitation in nuclear medicine AI applications due to the time-consuming, costly nature of dynamic PET acquisitions and lack of widespread availability (Rahmim & Zaidi, 2008; Townsend, 2020). Large patient cohorts with larger heterogeneity are required to address inter-patient heterogeneity in tracer kinetics, which is essential for clinical translation. Multi-institution collaborations and data sharing, i.e., The Cancer Imaging Archive (Clark et al., 2013), are the most significant opportunities to scale up future work.

The second limitation is that this research has considered only F-18 NaF tracer kinetics in prostate cancer bone metastasis. Though NaF PET is highly sensitive for skeletal lesion identification, generalization to other tracers (FDG, FMISO, or PSMA) and other cancers needs to be investigated. Further, though model predictions were contrasted with biexponential fitting, different kinetic models (two-tissue or multi-compartmental models) could have provided additional information regarding tracer behavior (Karakatsanis et al., 2013; Muzic & Cornelius, 2001). Having such models in the next studies would promote biological interpretability and facilitate direct comparison between AI-based prediction and traditional kinetic modeling paradigms.

4. Conclusions

The Long Short-Term Memory (LSTM) model demonstrates superior performance compared to the Recurrent Neural Network (RNN) in predicting the uptake patterns of the F-18 NaF radiopharmaceutical in the bone tissue of prostate cancer patients based on sequential PET/CT data. This superiority is reflected in higher determination coefficient (R^2) values (0.9128 compared to 0.9071) and lower prediction error values, both in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Additionally, LSTM demonstrates superior capability in stable and accurate representing biological kinetic dynamics, making it a more suitable architecture for predictive applications in temporal modeling of molecular imaging data.

Future work must address several shortcomings that remain in the current research. One important direction is the development of large multicenter cohorts, in which multiple institutions collect PET/CT data to capture greater variability in disease presentation and patient populations. Another area of focus is cross-tracer generalization, where determining whether FDG- or other tracer-trained models can effectively generalize to NaF imaging would significantly enhance the utility of such approaches. In addition, integrating predictive models into clinical decision support systems (CDSS) could provide real-time assistance to nuclear medicine physicians, thereby improving workflow efficiency. Advancing model performance will also require the identification and application of next-generation architectures, including transformers, graph neural networks, and CNN-RNN hybrids. Finally, clinical trials will be essential to evaluate whether AI-augmented kinetic prediction ultimately benefits patients and contributes to cost savings in real-world healthcare settings.

Acknowledgement

The authors would like to express their sincere gratitude to all parties who contributed to the completion of this research.

Author Contribution

All authors contributed equally to the conception, design, analysis, and writing of this manuscript.

Funding

This research received no external funding.

Ethical Review Board Statement

Not available.

Informed Consent Statement

Not available.

Data Availability Statement

Not available.

Conflicts of Interest

The authors declare no conflict of interest.

Open Access

©2025. The author(s). This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit: <http://creativecommons.org/licenses/by/4.0/>

References

- Carson, R. E., & Feng, D. (2021). Compartmental modeling in dynamic PET: Limitations and challenges in clinical translation. *Annals of Nuclear Medicine*, 35(8), 833–844. <https://doi.org/10.1007/s12149-021-01664-5>
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6), 1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
- Coleman, R. E. (2006). Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clinical Cancer Research*, 12(20), 6243s–6249s. <https://doi.org/10.1158/1078-0432.CCR-06-0931>
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30(9), 1323–1341. <https://doi.org/10.1016/j.mri.2012.05.001>
- Grant, F. D., Fahey, F. H., Packard, A. B., Davis, R. T., Alavi, A., & Treves, S. T. (2008). Skeletal PET with 18F-NaF: Applying new technology to an old tracer. *Journal of Nuclear Medicine*, 49(1), 68–78. <https://doi.org/10.2967/jnumed.107.045328>

- Gunn, R. N., Gunn, S. R., & Cunningham, V. J. (2001). Positron emission tomography compartmental models. *Journal of Cerebral Blood Flow & Metabolism*, 21(6), 635–652. <https://doi.org/10.1097/00004647-200106000-00002>
- Herzog, H., Lassen, M. L., & Ahn, S. (2020). Reducing scan duration in dynamic PET using AI-based reconstruction of missing time frames. *European Journal of Nuclear Medicine and Molecular Imaging*, 47(13), 3051–3061. <https://doi.org/10.1007/s00259-020-04888-7>
- Hospodarskyy, O., Martsenyuk, V., Kukharska, N., Hospodarskyy, A., & Sverstiuk, S. (2024, June). Understanding the Adam Optimization Algorithm in Machine Learning. *2nd International Workshop on Computer Information Technologies in Industry 4.0*. <https://doi.org/10.1016/j.ejomp.2023.103192>
- Iagaru, A., & Mittra, E. (2012). Advances in PET detection of bone metastases in prostate cancer. *Seminars in Nuclear Medicine*, 42(6), 465–480. <https://doi.org/10.1053/j.semnuclmed.2012.07.005>
- Ivashchenko, O. V., O'Doherty, J., Hardiansyah, D., Cremonesi, M., Tran-Gia, J., Hippeläinen, E., Stokke, C., Grassi, E., Sandström, M., & Glatting, G. (2024). Time-activity data fitting in molecular radiotherapy: Methodology and pitfalls. *Physica Medica*, 117, 103192. <https://doi.org/10.1016/j.ejomp.2023.103192>
- Jin, J. K., Dayyani, F., & Gallick, G. E. (2011). Steps in prostate cancer progression that led to bone metastasis. *International Journal of Cancer*, 128(11), 2545–2561. <https://doi.org/10.1002/ijc.26024>
- Kaalep, A., Sera, T., Rijnsdorp, S., Yaqub, M., Tohka, J., Boellaard, R., & Simplified Kinetic Modeling Taskforce. (2021). Variability and reproducibility of compartmental model outcome parameters in dynamic PET. *EJNMMI Physics*, 8(1), 77. <https://doi.org/10.1186/s40658-019-0265-8>
- Karakatsanis, N. A., Lodge, M. A., Tahari, A. K., Zhou, Y., Wahl, R. L., & Rahmim, A. (2013). Dynamic whole-body PET parametric imaging: Concept, acquisition protocol optimization and clinical application. *Physics in Medicine and Biology*, 58(20), 7391–7418. <https://doi.org/10.1088/0031-9155/58/20/7391>
- Kinahan, P. E., & Fletcher, J. W. (2010). Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Seminars in Ultrasound, CT and MRI*, 31(6), 496–505. <https://doi.org/10.1053/j.sult.2010.10.001>
- Kurdziel, K. A., Apolo, A. B., Lindenberg, L., Mena, E., McKinney, Y. Y., Adler, S. S., Turkbey, B., Dahut, W., Gulley, J. L., Madan, R. A., Landgren, O., & Choyke, P. L. (2015). *Data From NaF PROSTATE (Version 1) [Data set]*. The Cancer Imaging Archive. <https://doi.org/https://doi.org/10.7937/K9/TCIA.2015.ISOQTHKO>
- Kurdziel, K. A., Shih, J. H., Apolo, A. B., Lindenberg, L., Mena, E., McKinney, Y. Y., Adler, S. S., Turkbey, B., Dahut, W., Gulley, J. L., Madan, R. A., Landgren, O., & Choyke, P. L. (2012). The kinetics and reproducibility of 18F-sodium fluoride for oncology using current PET camera technology. *Journal of Nuclear Medicine*, 53(8), 1175–1184. <https://doi.org/10.2967/jnumed.111.100883>
- Lei, Y., Dong, X., Higgins, K., Liu, T., Curran, W. J., Mao, H., & Yang, X. (2022). Deep learning for dynamic PET parametric imaging using recurrent networks. *Physics in Medicine & Biology*, 67(6), 065002. <https://doi.org/10.1088/1361-6560/ac4b6d>
- Liu, G., Li, Y., Zhang, X., Chen, W., & Cherry, S. R. (2023). Total-body PET kinetic modeling: Opportunities and challenges for multi-organ quantification. *European Journal of Nuclear Medicine and Molecular Imaging*, 50(9), 2345–2358. <https://doi.org/10.1007/s00259-023-06299-w>
- Muraina, & Olaniyi, I. (2022). Ideal Dataset Splitting Ratios In Machine Learning Algorithms: General Concerns For Data Scientists And Data Analysts. *7th International Mardin Artuklu Scientific Researches Conference*, 496–504. <https://www.researchgate.net/publication/358284895>

- Muzi, M., Mankoff, D. A., Grierson, J. R., Wells, J. M., Vesselle, H., & Krohn, K. A. (2012). Kinetic modeling of ¹⁸F-NaF PET images of bone metastases. *Journal of Nuclear Medicine*, 53(12), 1904–1910. <https://doi.org/10.2967/jnumed.112.105080>
- Muzic, R. F., & Cornelius, S. (2001). COMKAT: Compartment model kinetic analysis tool. *Journal of Nuclear Medicine*, 42(4), 636–645. <https://jnm.snmjournals.org/content/42/4/636>
- Muzic, R. F., Lodge, M. A., & Rahmim, A. (2025). Advances in PET tracer kinetic modeling in the era of total-body scanners. *Annals of Nuclear Medicine*, 39(2), 215–227. <https://doi.org/10.1007/s12149-025-02014-x>
- Rahmim, A., & Zaidi, H. (2008). PET versus SPECT: Strengths, limitations and challenges. *Nuclear Medicine Communications*, 29(3), 193–207. <https://doi.org/10.1097/MNM.0b013e3282f3a515>
- Rajkomar, A., Dean, J., & Kohane, I. (2018). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMr1814259>
- Schwartz, J., Humm, J. L., Gonen, M., Kalaigian, H., Schoder, H., Larson, S. M., & Nehmeh, S. A. (2011). Repeatability of SUV measurements in serial PET. *Medical Physics*, 38(5), 2629–2638. <https://doi.org/10.1118/1.3578604>
- Shen, W., Zhou, M., Yang, F., & Yang, C. (2020). Predicting ¹⁸F-FDG PET dynamic images using deep learning for dose reduction. *Physics in Medicine & Biology*, 65(11), 115005. <https://doi.org/10.1088/1361-6560/ab896d>
- Tang, J., & Rahmim, A. (2021). Emerging approaches beyond compartmental models: Machine learning for PET kinetic analysis. *International Journal of Imaging Systems and Technology*, 31(3), 1122–1135. <https://doi.org/10.1002/ird3.66>
- Thie, J. A. (2004). Understanding the Standardized Uptake Value, Its Methods, and Implications for Usage. *Journal of Nuclear Medicine*, 45(9), 1431–1434. <https://jnm.snmjournals.org/content/45/9/1431>
- Townsend, D. W. (2020). Positron emission tomography/computed tomography. *Seminars in Nuclear Medicine*, 50(2), 123–131. <https://doi.org/10.1053/j.semnuclmed.2019.11.003>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>
- Wang, G., Qi, J., & Chen, K. (2021). Deep learning for PET parametric imaging: Accurate kinetic modeling from reduced acquisition protocols. *Medical Physics*, 48(6), 3204–3214. <https://doi.org/10.1002/mp.14873>

Biographies of Authors

Sani Inuwa Sani, Department of Geography, Mathematics and Natural Sciences, Universitas Indonesia, Depok, West Java 16424, Indonesia.

- Email: saniinuwa@gmail.com
- ORCID: 0009-0006-5136-1529
- Web of Science ResearcherID: NKQ-1969-2025
- Scopus Author ID: N/A
- Homepage: N/A

Mohammad Sidik Cahyana, Department of Medical Physics, Mathematics and Natural Sciences, Universitas Indonesia, Depok, West Java 16424, Indonesia.

- Email: mchsidiq12@gmail.com
- ORCID: 0009-0007-0734-0533
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A

Haryanto, Department of Physics, Mathematics and Natural Sciences, Universitas Indonesia, Depok, West Java 16424, Indonesia.

- Email: haryanto2409@gmail.com
- ORCID: 0009-0005-9553-7031
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A