

---

---

# Intelligent Imaging: Developing a Machine Learning Project

Geoffrey M. Currie

*School of Dentistry and Health Sciences, Charles Sturt University, Wagga Wagga, Australia; Department of Radiology, Baylor College of Medicine, Houston, Texas*

---

Artificial intelligence (AI) has rapidly progressed, with exciting opportunities that drive enthusiasm for significant projects. A sensible and sustainable approach would be to start building an AI footprint with smaller, machine learning (ML)-based initiatives using artificial neural networks before progressing to more complex deep learning (DL) approaches using convolutional neural networks. Several strategies and examples of entry-level projects are outlined, including mock potential projects using convolutional neural networks toward which we can progress. The examples provide a narrow snapshot of potential applications designed to inspire readers to think outside the box at problem solving using AI and ML. The simple and resource-light ML approaches are ideal for problem solving, are accessible starting points for developing an institutional AI program, and provide solutions that can have a significant and immediate impact on practice. A logical approach would be to use ML to examine the problem and identify among the broader ML projects which problems are most likely to benefit from a DL approach.

**Key Words:** nuclear medicine; artificial neural network; machine learning; artificial intelligence

**J Nucl Med Technol 2021; 49:44–48**

DOI: 10.2967/jnmt.120.256628

---

**A**rtificial intelligence (AI) has moved at such a fast pace and offers so many exciting opportunities that the inclination is to jump into something big. But a more sensible and sustainable approach would be to start small, learn, build momentum, and use internal knowledge to then grow the AI strategy. Individuals or groups eager to build their AI footprint tend to target their efforts to create a world-changing deep learning (DL) algorithm. Getting started in AI is not the real challenge; it is getting started without bursting the enthusiasm bubble. For example, consider a keen group of collaborators who want to develop a DL algorithm to stratify asthma and silicosis on ventilation lung scans from disease-free scans but with no AI experience, limited funding, and, to be fair, little insight into whether the lung ven-

tilation study has characteristics capable of differentiating the pathologies from normality. A lot of time, energy, and money could be wasted before the project failed. Even if the DL provided some stratification, humans would not have learned anything because we do not understand what is happening in the “magic box.” The first step is to determine whether there is a problem and whether AI has the capacity to be part of the solution. The better approach would use the wealth of information on a large patient series with known outcomes, including radiomic features extracted from the lung scan, and feed those features into a machine learning (ML)-based artificial neural network (ANN) and train it on the basis of a binary grounded truth of disease of interest versus no disease of interest. In doing so, we would allow the ANN to identify specific inputs or features that individually or collectively provide the highest predictive capability for the disease. The ANN may reveal that the lung scan plays an important role or no role at all. In the case of the former, potential to further automate through DL and a convolutional neural network (CNN) might be justified. In the case of the latter, we would avoid the wasted time and money associated with a DL approach. Those considering projects to develop a footprint in AI should keep in mind the following: consider a problem where the solution provides a good outcome rapidly and use that to build momentum (early win); avoid taking on a high-investment project (in time, resources, energy, and funding) that provides no reward for a long time; ensure that the outcome is not trivial; address a problem that has an impact and will generate attention with success; focus on projects that tangibly connect to nuclear medicine rather than peripheral areas of interest; ensure that your team comprises credible colleagues (internally and externally); and recruit a team that has complementary skills, is enthusiastic, and can accelerate both the project and the resulting AI momentum.

Developing an AI strategy for your research team or clinical department is not an overnight journey. Do not expect the exciting project convolved today to have clinical impact any time soon. Consider the scenario in which a data-rich department trains an ANN or CNN to perform a task and then validates that algorithm (1–3). This scenario represents almost all of the current literature on ML and DL in nuclear medicine. These are internally valid data that, with success, provide an algorithm that can reliably perform the prescribed task to enhance internal processes and outcomes.

---

Received Sep. 8, 2020; revision accepted Nov. 9, 2020.  
For correspondence and reprints contact: Geoffrey M. Currie, Charles Sturt University, P.O. Box U102, CSU, Wagga Wagga, NS 2650, Australia.  
E-mail: gcurrie@csu.edu.au  
Published online Dec. 24, 2020.  
COPYRIGHT © 2021 by the Society of Nuclear Medicine and Molecular Imaging.

Commercialization of these algorithms for a broader impact is much more difficult. First, the process of data entry, sharing, and management needs to overcome issues of privacy and security (4). Second, the internal dataset may contain local bias that threatens external validity, thus requiring rigorous validation with external data. Third, the algorithm may be specific for internal protocols and equipment and may not hold true when parameters or equipment change. For example, the ANN scaling and weighting associated with predicting cardiac events in patients with  $^{123}\text{I}$ -metaiodobenzylguanidine heart failure (5) change substantially if the uptake phase is undertaken in temperatures that vary by as little as  $2^\circ\text{C}$ , if the time from injection to scanning time varies, if the collimator or energy window varies, and if the specific  $\gamma$ -cameras have different resolutions or sensitivities. Beyond these variations, changing the acquisition or reconstruction parameters will also change scaling and weights. These issues provide insight into the discrepancy between the enormous potential applications of ML and DL in nuclear medicine and radiology and the actual number of commercially available algorithms approved by the U.S. Food and Drug Administration.

Consideration should be given to how the ML output can be neatly integrated into the existing graphical output of nuclear medicine scans. This, in part, also speaks to the commercializability of the algorithm. The addition of ML-based risk stratification and scoring is a logical step. Consider the left side of Figure 1 from the context of the feasibility and clinical utility of a CNN-driven risk score for pulmonary embolism that can be summarized and displayed in a simple format. The right side of Figure 1 is perhaps an option in which the perfusion SPECT is segmented against the accompanying low-dose CT to predict pulmonary embolism. Segmentation, risk scoring of individual lesions, and determining total disease burden might be helpful for patients who are being evaluated for metastatic spread to bone (Fig. 2). These are all mock-ups for potentially useful AI projects to provoke your own thoughts and ideas. These examples should not constrain your vision, with innovative applications of AI falling well beyond these simple examples including classification, localization, detection and segmentation projects.

## ML EXAMPLES

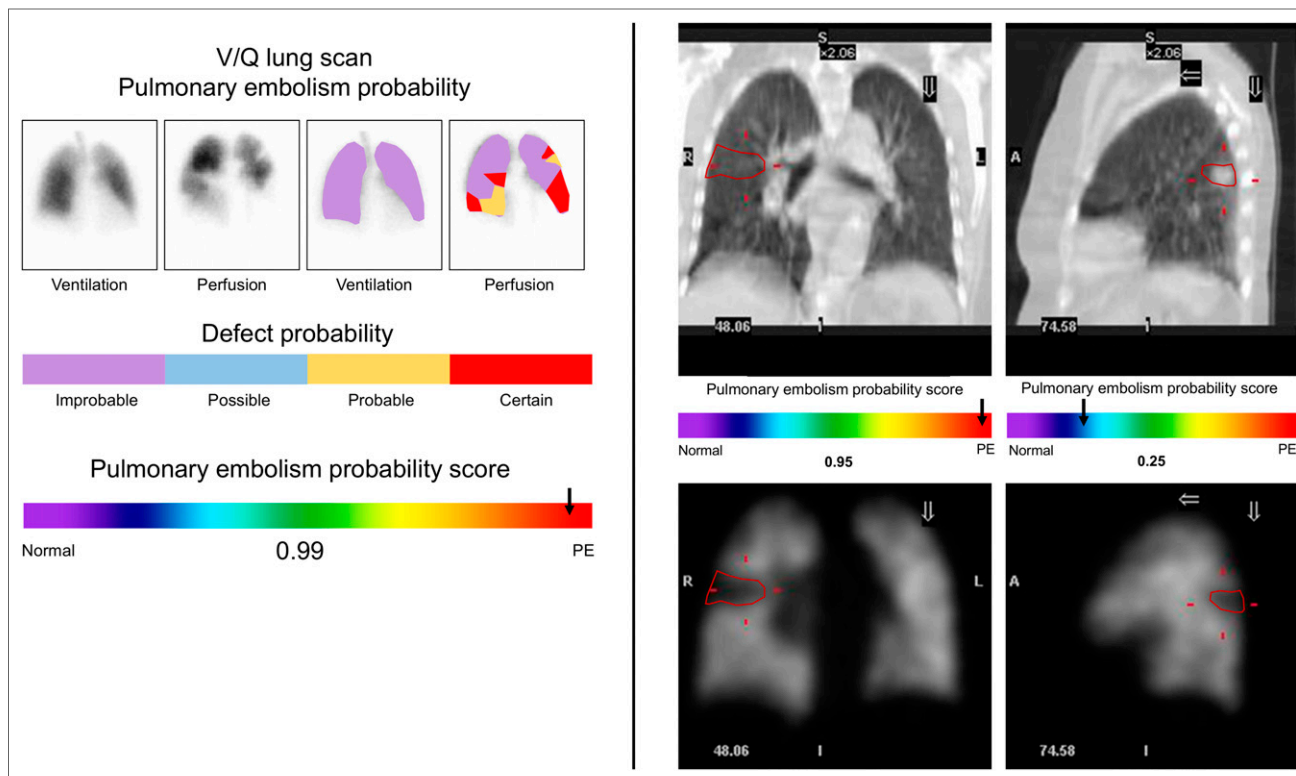
The following are 2 simple ML projects using an ANN. This kind of approach can be undertaken with custom development tools or using commercial software packages. In one project, the key predictor of a cardiac event in heart failure patients was identified to be regional analysis of myocardium immediately adjacent to infarcted tissue and the associated  $^{123}\text{I}$ -metaiodobenzylguanidine washout (6). This regional SPECT analysis was to provide richer insights than the previously described global approach to planar analysis based on the fact that the pathology cascades from regional to global. The conventional statistical analysis for the out-

come was complex and individualized but provided the greatest single predictor of a cardiac event and, thus, the need for an implantable cardioverter defibrillator.

The emergence of capability in developing ANNs and their use as tools for data analysis in parallel to conventional statistical analysis saw the data reanalyzed. The advantage of the ANN is that it examines all variables in a manner that removes duplication (even if highly correlated) and redundancy, weights variables to optimize the predictive outcome, and, in doing so, identifies the variable or combination of variables that provides the greatest predictive power. The ANN identified the same single best predictor as did conventional statistical analysis (washout from tissue immediately adjacent to the infarcted tissue) but also determined that the combination of 2 very simple variables provided the highest overall predictive capability (Fig. 3, top) (6). In this case, the ANN was not DL or CNN but rather an ML tool designed to analyze more critically the relationships between variables and outcome (cardiac event) prediction.

The ANN architecture initially comprised 84 scaling layer inputs—3 hidden layers of 6 nodes each using a logistic activation function for the binary output layer. The weighted-squared-error method was used to determine the loss index, and the neural parameters norm was used for the regularization method. A quasi-Newton training method was adopted, and the architecture was optimized at 2 inputs (change in left ventricular ejection fraction and planar washout), 2 hidden layers of 6 and 1 node, and a binary output. This architecture was evaluated using receiver-operator-characteristic analysis, which demonstrated an area under the curve of 0.75, correlating to a sensitivity of 100% and a specificity of 50%. The ANN indicated that the best predictor of a cardiac event is the coexistence of a decrease in the left ventricular ejection fraction of 10% or greater and a planar washout of 30% or greater.

A second example was part of a detailed analysis of metrics to characterize injection kinetics in PET using topical sensors (7). The commercial software provided several metrics and scores for interpretation by users. Although individual metrics provide useful tools, combining the errors associated with each produced an overall larger error. There were 45 input variables in 863 patients, using a binary classification of dose extravasation or no dose extravasation. The heat-map/correlation matrix identified several redundant variables, and the highest correlation scores were associated with tc50 (0.838), ndAvgN (0.749), and CEnd ratio (0.721), consistent with the conventional statistical analysis. The tc50 is the metric that represents the point in time where the injection sensor counts revert to being double that of the reference sensor, ndAvgN is the difference between injection and reference sensors at 4 min after injection, and CEnd ratio is the ratio of sensor to reference sensor counts at the end of the monitoring period (45–60 min). A 60:20:20 split was used for training, selection, and testing. The ANN architecture included 3 hidden layers of



**FIGURE 1.** On left is mock summary output for CNN-based risk algorithm for pulmonary embolism using ventilation and perfusion mismatch. On right is mock summary output for CNN-based risk algorithm for pulmonary embolism using low-dose CT and perfusion SPECT mismatch. Coronal and sagittal slices represent 2 different patients: one with mismatch consistent with high likelihood of pulmonary embolism (coronal), and another with matching defect associated with lower likelihood of pulmonary embolism (sagittal). PE = pulmonary embolism; V/Q = ventilation-perfusion.

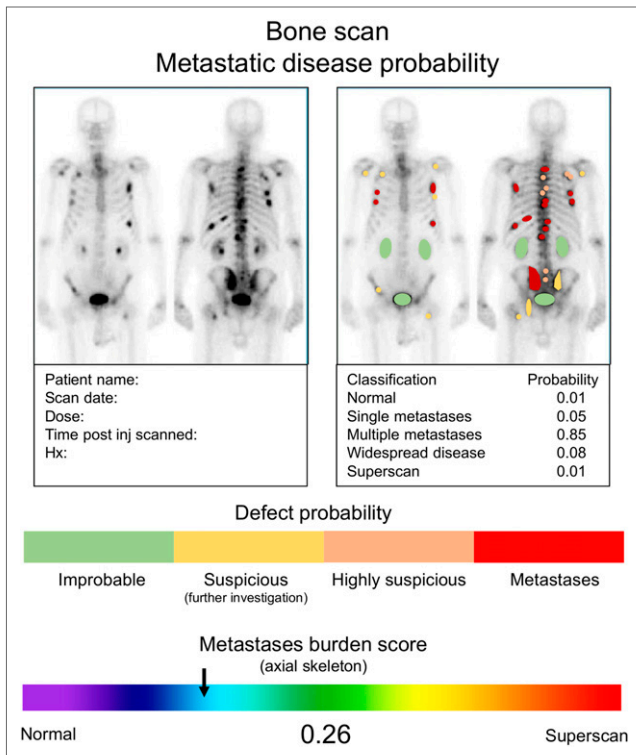
5, 5, and 1 nodes, using a logistic activation function for a binary output. As with the first example, the weighted-squared-error method determined the loss index, the neural parameter norm was applied for the regularization method, and a quasi-Newton training method was used. The final architecture of the ANN reflects the optimized subset of inputs with the lowest selection loss, in this case, 4 inputs; 3 hidden layers of 3, 4, and 1 node; and a binary output (Fig. 3, bottom). The purpose of removing redundancy and optimizing both the number of inputs and the complexity (number and depth of nodes) is to minimize over- or underfitting. This was evaluated with receiver-operator-characteristic analysis, which demonstrated an area under the curve of 1.0, correlating with a sensitivity of 100% and a specificity of 100% although binary classification testing revealed a less than perfect predictive model, with classification accuracy of 0.75.

#### CODE, FRAMEWORKS, AND LIBRARIES

The tools for AI (Fig. 4) can be generally classified as datasets (serving the appetite of DL for data), architecture (functional model), framework (to enhance functionality), hardware (e.g., central processing unit and graphics processing unit), libraries (sources of datasets, architectures, and frameworks), and code (language). Python (<https://www.python.org/doc/>) is a high-level open-source, fairly

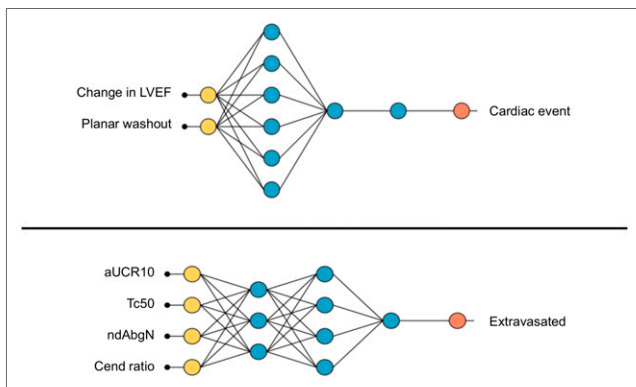
simple, and intuitive programming language that can be used for direct coding in ML and DL or the source code for Keras-based network development (application training interface). It is not the only programming language used for CNN development but is perhaps the most common and certainly the most accessible and usable for those early developers. NumPy (<https://numpy.org/>) is an extension package for Python that allows algebraic functions, Fourier transformation, and integration of Python code with other types of code (Fortran and C, for example). NumPy is like adding the scientific calculator to the standard calculator, enhancing Python's scientific functionality.

Keras (<https://keras.io/>) is a DL library that runs as an application programming interface for Python in conjunction with TensorFlow (<https://www.tensorflow.org/>). PyTorch (<https://pytorch.org/>) is an alternative open-source library of tools for training-phase platforms. TensorFlow is a graphics-based computation platform that stores data in objects (tensors) rather than integers or strings. It is used as a back-end platform for CNN deployment but can include Keras for broader application in the training phase, providing an end-to-end platform. The framework and associated libraries are described as the ecosystem tools, whereas the architecture and datasets are the landscape. Caffe (<https://caffe.berkeleyvision.org/>) is the University of California,

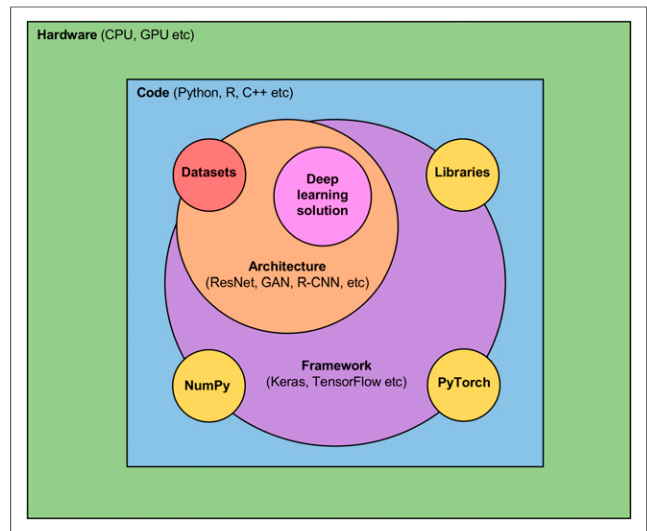


**FIGURE 2.** Mock summary output for CNN-based risk algorithm for skeletal metastases, with probability classification for various outcomes and risk assessment for individual lesions. Hx = history; post inj = after injection.

Berkley, DL framework and is useful for those a little more adventurous with DL research. KNIME (<https://www.knime.com/>) is a DL analytics platform for CNN development, training, and implementation. KNIME integrates with the Keras DL architecture and the TensorFlow ML library. Theano (<https://github.com/Theano/Theano>) is a similar library for scientific extension of Python as NumPy, with additional features



**FIGURE 3.** Final ANN architecture optimizes inputs and network complexity for heart failure cohort (top) and dose extravasation cohort (bottom). Here, number of inputs, and number and complexity of nodes, are optimized. aUCR10 = area under the curve ratio between 1 and 10 min; Cend = counts at the end of the data collection; LVEF = left ventricular ejection fraction; ndAbgN = difference in counts at 4 min; Tc50 = time until reference counts are half of sensor counts.



**FIGURE 4.** Schematic representation of relationships between hardware, code, frameworks, architectures, libraries, datasets, and DL solution. DL landscape is defined by 2 orange zones, and DL ecosystem, by purple and yellow zones. CPU = central processing unit; GAN = generative adversarial network; GPU = graphics processing unit; R-CNN = region-based convolutional neural network; ResNet = residual neural network.

associated with speed and optimization. Deeplearning4j (<https://deeplearning4j.org/>) is a Java-based, open-source DL library that can be integrated with Keras as the Python application programming interface. Pandas (<https://pandas.pydata.org/>) are powerful data analysis toolkits for Python, essentially a library of tools for those using Python for data analysis.

### CONVOLUTION NEURAL NETWORKS

The CNN provides a rich range of opportunities for development of AI projects. These include CNNs for automatic segmentation of complex datasets, disease detection and segmentation, image reconstruction methods, development of pseudo-CT for attenuation correction of PET/MRI or PET alone, scatter correction and denoising of images, and radiation dosimetry on serial PET images. These approaches could adopt a variety of CNN architectures (e.g., AlexNet or VGGNet), encoder/decoder architectures (e.g., UNet), and generative adversarial networks. An entry-level CNN project might avoid the complexity of dynamic data or tomographic data. Instead, a simple CNN architecture (e.g., AlexNet) could be applied either using direct python coding or via commercial software (e.g., DL toolbox from MATLAB; <https://au.mathworks.com/products/deep-learning-hdl.html>) to run a simple classification project on planar images. Consider a simple import of a lung perfusion scan into a CNN trained against known cases of pulmonary embolism or otherwise. The CNN, once trained and validated, could triage perfusion scans with findings consistent with pulmonary embolism for more urgent reporting. This avoids the greater complexity of comparing ventilation studies, because

the role is to triage, not to diagnose (e.g., the graphical abstract).

## CONCLUSION

AI and ML have almost infinite applications in nuclear medicine and medical imaging. The examples above are far from exhaustive. They provide a narrow snapshot of potential applications designed to inspire readers to think outside the box at problem solving using AI and ML. The key consideration is identifying a problem whose solution would bring tangible benefits and identifying the appropriate tool to solve it. In many cases, solving a problem does not need AI, ML, or DL. Frequently, the simpler and less resource-intensive ML approaches are ideal for problem solving. Indeed, the plethora of ML-based applications in nuclear medicine and medical imaging not only are accessible starting points for a developing institutional AI program but also provide solutions that can have a significant and immediate impact on practice and patient outcomes. Nonetheless, ML applications are often overlooked in favor of the higher-order, perhaps more “prestigious” DL approaches that demand higher levels of expertise, time, and resources. A logical approach would be to use

ML to examine the problem and identify, among the broader ML projects, which problems are most likely to benefit from a DL approach.

## DISCLOSURE

No potential conflict of interest relevant to this article was reported.

## REFERENCES

1. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci.* 2019;50:477–487.
2. Currie G. Intelligent imaging: anatomy of machine learning and deep learning. *J Nucl Med Technol.* 2019;47:273–281.
3. Currie GM. Intelligent imaging: artificial intelligence augmented nuclear medicine. *J Nucl Med Technol.* 2019;47:217–222.
4. Currie G, Hawk KE, Rohren E. Ethical principles for the application of artificial intelligence (AI) in nuclear medicine and molecular imaging. *Eur J Nucl Med Mol Imaging.* 2020;47:748–752.
5. Iqbal B, Currie G, Kiat H. Regional cardiac sympathetic nervous system evaluation using  $^{123}\text{I}$ -mIBG SPECT in patients with heart failure. *J Med Imaging Radiat Sci.* 2018;49:397–405.
6. Currie G, Iqbal B, Kiat H. Intelligent imaging: radiomics and artificial neural networks in heart failure. *J Med Imaging Radiat Sci.* 2019;50:571–574.
7. Currie GM, Sanchez S. Topical sensor metrics for  $^{18}\text{F}$ -FDG positron emission tomography dose extravasation. *Radiography.* 2021;27:178–186.