AI-Based Image Quality Assessment in CT

Lars Edenbrandt¹,², Elin Tragardh⁴,⁵, and Johannes Ulen⁶

Abstract
Medical imaging, especially computed tomography (CT), is becoming increasingly important in research studies and clinical trials and adequate image quality is essential for reliable results. The aim of this study was to develop an artificial intelligence (AI)-based method for quality assessment of CT studies, both regarding the parts of the body included (i.e. head, chest, abdomen, pelvis), and other image features (i.e. presence of hip prosthesis, intravenous contrast and oral contrast).

Approach: 1,000 CT studies from eight different publicly available CT databases were retrospectively included. The full dataset was randomly divided into a training (n = 500), a validation/tuning (n = 250), and a testing set (n = 250). All studies were manually classified by an imaging specialist. A deep neural network network was then trained to directly classify the 7 different properties of the image.

Results: The classification results on the 250 test CT studies showed accuracy for the anatomical regions and presence of hip prosthesis in the interval 98.4% to 100.0%. The accuracy for intravenous contrast was 89.6% and for oral contrast 82.4%.

Conclusions: We have shown that it is feasible to develop an AI-based method to automatically perform a quality assessment regarding if correct body parts are included in CT scans, with a very high accuracy.

Keywords: Artificial Intelligence; Diagnostic Imaging; Machine Learning; Quality Assessment

Introduction
Medical imaging, especially computed tomography (CT), is becoming increasingly important in research studies and clinical trials. Large projects and trials could include hundreds or thousands of CT studies and adequate image quality is essential for reliable results. This is a particular concern in multi-center trials, which often provide detailed imaging guides that must be followed in order to correctly include patients. Problems related to imaging could lead to either exclusion of patients or false image data incorporated in study or trial results. A quality check of images selected for a study is therefore an important process. Today, this is performed manually. Often, the quality check must be performed promptly by the clinical research organization before the patient can be enrolled. In both clinical trials and in large retrospective studies, this could be tedious work. The quality check of images could ensure for example that:

• The correct part of the body is visible in the CT, e.g., the chest in a lung cancer study,

Affiliation:
¹Region Vastra Gotaland, Sahlgrenska University Hospital, Department of Clinical Physiology, Gothenburg, Sweden
²Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
³SliceVault AB, Malmo, Sweden
⁴Department of Clinical Physiology and Nuclear Medicine, Skane University Hospital and Lund University, Malmo, Sweden
⁵Wallenberg Center for Molecular Medicine, Lund University, Malmo, Sweden
⁶Eigenvision AB, Malmo, Sweden.

*Corresponding author:
Lars Edenbrandt, Region Vastra Gotaland, Sahlgrenska University Hospital, Department of Clinical Physiology, Gothenburg, Sweden.

Citation: Lars Edenbrandt, Elin Tragardh, and Johannes Ulen. AI-Based Image Quality Assessment in CT. Archives of Clinical and Biomedical Research 6 (2022): 869-872.

Received: August 25, 2022
Accepted: September 06, 2022
Published: October 17, 2022
• CT artefacts are not present, e.g., hip prostheses causing artefacts which usually prevent a proper analysis of the prostate,
• The CT is acquired according to the study protocol regarding the use of intravenous or oral contrast.

Information about a CT study should ideally be described in the DICOM tags. However, experience shows that it is not possible to rely only on this information. This is true especially in research projects and clinical trials when important information in the DICOM tags could be deleted in the pseudonymization process. The use of artificial intelligence (AI) to solve clinical problems has been intensely studied in recent times [1]. Deep learning, in particular, has gained attention as a method of obtaining complex information from medical images. AI could potentially be trained to help with image quality assessment and could be an important tool in this important, otherwise manual, task. The aim of this study was to develop an AI-based method for quality assessment of CT studies, both regarding the parts of the body included (i.e. head, chest, abdomen, pelvis), and other image features (i.e. presence of hip prosthesis, intravenous contrast and oral contrast).

Methods

Patients

We retrospectively included 1,000 CT studies from eight publicly available CT databases, see Table 1.

Before training the model, the full image set of 1,000 CT studies was randomly divided into a training (n = 500), a validation/tuning (n = 250), and a testing set (n = 250). The test set was reserved for model evaluation.

Manual Classification/Ground Truth Definition

All CT studies were classified by a nuclear medicine specialist experienced in hybrid imaging. Each case was classified based on the presence of the following seven features:
• Head- The cranium is visible at least partly. Head is not present if only part of the mandible is visible.
• Chest- The lungs are visible. Only very minor parts may be missing.
• Abdomen- Main parts of liver, spleen, and the kidneys are visible.
• Pelvis- The hip bones are visible.
• Hip prosthesis- Uni- or bilateral hip prosthesis including implants for fixation of hip fractures.
• Intravenous (IV) contrast- Signs of intravenous contrast including different phases.
• Oral contrast- Signs of oral contrast including different phases.

An overview of the distribution of the different classes in the dataset is given in Table 2.

AI Tool

The AI tool consists of a 3D-ResNet, [17] a deep neural network designed for classification of 3D images. The network have an input shape of $110 \times 110 \times 110 \times 1$ pixels with 7 output channels each with its own sigmoid activation. Each output channel represents one of the classes defined in Section 2.2. Many CT images contain quite a lot of air, which is not helpful for classification. In order to remove air, the images are smoothed using a Gaussian kernel with standard deviation $5 \text{mm}^3$. An axis-aligned bounding box is then fitted to all pixels with Hounsfield unit (HU) above $–800$ in the smoothed image. The original image is then cropped to this bounding box. The cropped CT images are pre-processed by clamping the HU values to the range $[-1000, 3000]$ and then normalized to $[-1, 1]$. Furthermore, the CT volumes are re-sized to resolution $5 \times 6 \times 12 \text{mm}$ (or the smallest possible pixel shape with the same aspect ratio making the full image fit) and placed in the middle of the input volume.

Sampling: The classes are quite imbalanced as seen in Table 2. In order to sample uncommon examples more often each image $i$ is sampled proportional to a weight $w_i$ defined as:

$$w_i = \sum_{\ell \in L} \begin{cases} \frac{1}{p_i}, & \text{if } \ell \text{ is true for image } i, \\ \frac{1}{n_i}, & \text{if } \ell \text{ is false for image } i, \end{cases}$$  

Table 1: The number of images selected from each publicly available database.

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of images</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4KC-KiTS</td>
<td>204</td>
<td>[2–4]</td>
</tr>
<tr>
<td>ACRIN 6668</td>
<td>203</td>
<td>[2, 5, 6]</td>
</tr>
<tr>
<td>CT Lymph Nodes</td>
<td>176</td>
<td>[2, 7–9]</td>
</tr>
<tr>
<td>CT-ORG</td>
<td>117</td>
<td>[2, 10–12]</td>
</tr>
<tr>
<td>NSCLC-Radiomics</td>
<td>115</td>
<td>[2, 13, 14]</td>
</tr>
<tr>
<td>Task 07 Pancreas</td>
<td>100</td>
<td>[15]</td>
</tr>
<tr>
<td>Task 03 Liver</td>
<td>52</td>
<td>[15]</td>
</tr>
<tr>
<td>Anti-PD-1 Immunotherapy Lung</td>
<td>33</td>
<td>[2, 16]</td>
</tr>
<tr>
<td>Total</td>
<td>1,000</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Positive and negative examples for each class in the dataset.

<table>
<thead>
<tr>
<th>Class</th>
<th>Positive count</th>
<th>Negative count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>256</td>
<td>744</td>
</tr>
<tr>
<td>Chest</td>
<td>603</td>
<td>397</td>
</tr>
<tr>
<td>Abdomen</td>
<td>903</td>
<td>97</td>
</tr>
<tr>
<td>Pelvis</td>
<td>702</td>
<td>298</td>
</tr>
<tr>
<td>Hip prosthesis</td>
<td>25</td>
<td>975</td>
</tr>
<tr>
<td>Intravenous contrast</td>
<td>256</td>
<td>744</td>
</tr>
<tr>
<td>Oral contrast</td>
<td>422</td>
<td>578</td>
</tr>
</tbody>
</table>
where \( L \) is the set of labels, \( p_l \) the total number of positive examples, and \( n_l \) the total number of negative examples for label \( l \). \( p \) and \( n \) are calculated individually for the training and validation sets.

**Training:** Binary cross-entropy is used as loss function and the network is optimized using the ADAM optimizer [18] with Nesterov momentum and an initial learning rate of \( 1 \times 10^{-5} \). Each training and validation epoch consists of 2,000 samples and 400 samples respectively. If the validation loss has not improved for 10 epochs the learning rate is halved until it reaches a minimum value of \( 1 \times 10^{-8} \). The training stops when validation loss has not improved for 20 epochs. During training the images are augmented using rotations (−0.1 to 0.1 radians), scaling (−10 to 10%) and an intensity shift of (−100 to +100 HU).

**Results**

The classification results on the test set of is presented in Table 3. The accuracy for the anatomical regions and presence of hip prosthesis were 98.4% to 100.0%. The accuracy for intravenous contrast was 89.6% and for oral contrast 82.4%. Figure 1 and 2 show patient examples with correct and non-correct classifications.

**Discussion**

In this study we have shown that it is feasible to develop an AI-based tool to automatically check that the correct body parts are visible in the CT studies, with a very high accuracy. The AI-based method was also able to accurately detect hip prosthesis even though the number of positive cases in the training and validation sets were limited (\( n = 19 \)). A limitation of this study was that the AI-based tool only made a classification regarding presence of contrast or not. Many different phases of contrast enhancement exist, [19] typically early arterial phase (15–25 s post injection), late arterial phase (30–40 s post injection), hepatic or late portal venous phase (70–90 s post injection), nephrogenic phase (85–120 s post injection) and excretory or delayed phase (5–10 min post injection). No clearly defined times post injection of the contrast agent exist, but with a large number of images with different contrast phases in the training group, it would probably be possible to train an AI-method to categorize the contrast phase in more detail than we did in this study. Other potential problems related to intravenous contrast is different amounts of contrast agent administered, for example reduced doses in patients with kidney disease. Problems with oral contrast for this type of task include different timings of contrast as well as different contrast agents (for example barium or iodine-based agents). A more comprehensive classification of contrast would most likely require a much larger training set. Some of the false negative cases of our test set represented very late intravenous phases with low contrast in the aorta but contrast in the kidneys or urinary bladder (Figure 2). This type of cases was not common in the training set. Also the appearance of oral contrast on CT showed substantial variation. In most cases contrast was clearly visible in the small bowel. In other cases only the stomach or colon showed contrast. Medical imaging is often a key asset in clinical trials, as it can provide efficacy evaluation and safety monitoring [20]. It is also often used as screening for eligible patients to include. Medical imaging can also improve clinical trial efficacy and reduce the time to complete a specific trial, by offering imaging biomarkers that can act as a surrogate endpoint. In order to do so, good image quality is crucial and therefore it is necessary to monitor image quality throughout different stages of a study. A step in the quality assessment could be to determine if correct body

**Table 3:** Result for the 250 test CT studies. True positive (TP), True negative (TN), False positive (FP), False negative (FN), and accuracy.

<table>
<thead>
<tr>
<th>Classification task</th>
<th>TP</th>
<th>TN</th>
<th>FP</th>
<th>FN</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>70</td>
<td>176</td>
<td>0</td>
<td>4</td>
<td>98.40%</td>
</tr>
<tr>
<td>Chest</td>
<td>146</td>
<td>104</td>
<td>0</td>
<td>0</td>
<td>100.00%</td>
</tr>
<tr>
<td>Abdomen</td>
<td>222</td>
<td>24</td>
<td>0</td>
<td>4</td>
<td>98.40%</td>
</tr>
<tr>
<td>Pelvis</td>
<td>167</td>
<td>80</td>
<td>0</td>
<td>3</td>
<td>98.80%</td>
</tr>
<tr>
<td>Hip prosthesis</td>
<td>6</td>
<td>244</td>
<td>0</td>
<td>0</td>
<td>100.00%</td>
</tr>
<tr>
<td>Intravenous contrast</td>
<td>149</td>
<td>75</td>
<td>13</td>
<td>13</td>
<td>89.60%</td>
</tr>
<tr>
<td>Oral contrast</td>
<td>71</td>
<td>135</td>
<td>15</td>
<td>29</td>
<td>82.40%</td>
</tr>
</tbody>
</table>

Figure 1: Example of a correctly classified image from ref [5]. Head, chest, abdomen, pelvis, hip prosthesis, and oral contrast were present. Intravenous contrast was not present.

Figure 2: Example of image with a miss-classification from ref [5]. Excretory intravenous contrast phase is present, but not detected by AI. All other classifications were correct.
parts are included and if the images contain contrast agent or not. Further development of automated image quality assessment could also include image properties such as noise level and patient motion. Evaluation by a human observer is both time consuming and subjective. AI-based tools could help minimize both issues.

Conclusions
We have shown that it is feasible to develop an AI-based method to automatically perform a quality assessment regarding if correct body parts are included in CT scans, with a very high accuracy.

Acknowledgment
We would like to thank Måns Larsson and Olof Enqvist for fruitful discussions regarding this study and closely related topics.

References