The following literature review provides insight into the existing works involving the medical imaging and machine learning, dose-related noise in CT and how these have been combined. It begins by providing context in lung functional analysis with both deterministic and data-driven methods. The description, simulation, and removal of noise in X-Ray and Computed Tomography are then discussed. Gaps in this literature are then stated, and the relevant and previous end-to-end denoising and machine learning designs are reviewed.

Automatic Lung Functional Analysis is a well-established area of study. With the recent explosion of applied machine learning research, the area has seen significant progress. [3] is an excellent example of this. It analyses a few hundred papers published in 2021, that used machine learning to diagnose COVID-19. It focusses on investigating the clinical relevance of these papers, finding none of those provided to be clinically applicable. It goes on to provide common pitfalls and recommendations going forward for these kinds of problems, which will be particularly relevant in this thesis, in terms of how to go about developing, and assessing the model. The number of papers reviewed in [3] demonstrates how current and relevant the study area is, and thus how important it is to critically investigate the metrics and methods used in its development.

Generally, the approaches to this type of problem can be categorized into 1) deterministic and 2) data-driven methods. Here the term “deterministic” is used to refer to direct methods such as those in classical computer vision, where a given input image will always provide the same output, while “data-driven” refers to models involving machine learning with results based on training data. Earlier models, such as those described in [4] and [5] provide examples of deterministic models used for lung segmentation. While [5] is more focused on the segmentation step, [4] demonstrates the effectiveness on the end goal – the task-based accuracy. Without any machine learning techniques, and a very limited dataset ( only 38 scans), the method was quite successful