Master Thesis | Interdisciplinary M.Sc Course in Nanomedicine

National and Kapodistrian University of Athens

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens
— EST. 1837 —

Standardization of texture analysis

using IQ phantom

under preclinical PET imaging

Foivos Sokratis Lazaris

<u>**Date of Submission:**</u> **09/2020**

Academic Year 2019-2020

Foivos Sokratis Lazaris, B.Sc in Biology

Supervisor:

Anastasios Gaitanis, Associate Professor Level, BRFAA

Graduation Committee:

Efstathios Eftstathopoulos, Professor, University of Athens

Ellas Spyratou, M.Sc, Ph.D, University of Athens

Title: **"Standardization of texture analysis using IQ phantom under preclinical PET imaging".**

Abstract

Radiomics is a newly emerging field, claiming that there is a lot more information than a "naked" eye can conceive in medical images and needs to be extracted in the form of texture features. Medical images acquired through any imaging system (e.g. Positron Emission Tomography (PET), Computed Tomography (CT), Ultrasound (US), or Magnetic Resonance (MR)) are suitable candidates. However, a lot of research is required to identify the texture indices, which firstly, under certain conditions are reliable and secondly, can be associated with certain biological processes. Radiomics aim to find those indices and use them in diagnosis, prognosis, and treatment evaluation (theranostics). In this research work, we shall investigate, under preclinical conditions, the impact of bin size (total number of grey levels) in texture features and the repeatability, reproducibility, and liability of them (under specific conditions, both in terms of image reconstruction and acquisition time) using the NEMA NU 4-2008 image quality (IQ) phantom. For the acquisition of the PET images a small-animal PET/CT scanner – Mediso nanoScan®PC (PET8/2) – was employed. LIFEx version 6.20 (an open-source medical imaging analysis software) was incorporated for the analysis of PET images and the extraction of texture features (First-Order features, Second-Order features, and Higher-Order features). Statistical analysis (PD, COV, Interclass Correlation Coefficient, Spearman Rank Correlation Coefficient, one-way Analysis of Variance, and Least Significance Test) and evaluation of the indices was performed. After the evaluation with IQ phantom images, a preliminary study using PET image data from treated and non-treated mice was performed. The vast majority of the texture features showed low-to-no robustness and a high degree of variance, thus, making them bad candidates for potential biomarkers. Only a limited number of features found to perform well at all statistical tests (Histogram Skewness, Kurtosis, and Excess Kurtosis, GLCM Correlation, and NGLDM Coarseness), while, even less (Histogram Skewness and GLCM Correlation) could provide valuable information to distinguish the mice of the Treated from the Non-Treated group. Furthermore, we are proposing that the robustness of the features is directly associated with the total number of bins – First-Order and Second-Order features tend to be more robust at a higher number of bins, while Higher-Order features show greater robustness at a lower number of bins.

Table of Contents

1. Introduction

1.1 Thesis Originality & Issues Addressed

One of the main challenges regarding texture analysis is the need for the standardization of the protocols, since there is a huge heterogeneity at multiple levels (e.g. the process of the computation of the radiomic features, statistical analysis of the emerged data, protocols regarding image acquisition and reconstruction parameters) between different works. In the current work the impact of SUV discretization criteria (bin size – bin size is a custom number which determines the total discrete values that the software will be able to recognize in a given image (e.g. ΔSUV=0.5 MBq, bin size=0.05, total number of discrete values=10), the number of the discrete depends on the total ΔSUV as well) on the computation of the radiomic features is examined.

1.2 History of Medical Image Analysis

Origins of medical imaging are first found as early as the invention of X rays by a German professor of physics, Wilhelm Rontgen, in 1895. This resulted in a growing interest in the newly born field of Radiology, which quickly took off with the arrival of World War II. The first application of medical imaging, took place using X rays, which would pass through the body onto some film, generating an image. The whole process would take approximately 11 minutes (today's X rays imaging lasts just some milliseconds), while the patient was receiving an excessive dose of radiation – up to 50 times more than an X ray today [1].

A lot of improvements were made since this very first attempt, especially during the second half of the 20th century, but the one thing, which revolutionized the field of medical imaging was the invention on Ultrasound scanning (US) – which is closely related with the use of sonar, during World War II. It allowed the generation of useful and

accurate images – which contain medical valuable information – without exposing the patient to harmful radiation. A transducer sends a stream of high frequency, low wavelength soundwaves penetrating and travelling through the body's soft and hard tissues, bouncing back and echoes when hitting denser surfaces (e.g. organs and bones). US is used extensively today in clinical practice – ranging from the evaluation of fetal development to pregnancy progress – because of its safety, while it may also assist in some types of biopsy [1,2].

Another milestone in the development of the modern medical imaging took place in the 1970s when digital imaging techniques became widely available. Godfrey Hounsfield was the first to introduce computed tomography (CT) in 1972. Thus, allowing: the generation of higher quality images with a smaller radiation dose, the ability to enhance images using special contrast agents or post-acquisition processing (with software incorporation), and improved storage of records – while it allowed the development of the first databases (containing a huge amount of medical images) and the extraction of information that were not available before. In 1973, Haralick et al. [36] introduced the image texture analysis in an attempt to distinguish aerial from satellite photographs and acquire valuable information based on texture. Years later, with the improvements in texture analysis and the processing of the digital images, new medical subfields (Radiomics, Texture Analysis, Precision Medicine) will come up which will enable the extensive analysis of the medical images and the development of predictive biomarkers based on the image's texture. Hounsfield's original CT scan was using X rays and a detector mounted on a special rotating frame paired with a digital computer to produce in-detail cross sectional images of objects and would approximately last for a couple of hours just to acquire a single slice of image and up to 24 hours to reconstruct – the reconstruction algorithms were based on the theories of the mathematician Johan Radon [53], published in 1917 – this data into a single image (today's CT has the ability to acquire slices of an image in a matter of seconds, while the reconstruction happens immediately) [1,3,4].

Figure 1. Display of the first CT scanner developed by Godfrey Hounsfiled [63].

Even though magnetic resonance (MR) principles were known from the 1950s – different materials are able to resonate at different magnetic field strengths – as a product of the nuclear medicine project, which occurred from the 1940s till the 1990s and was funded by Development of Energy (DOE) Office of Biological and Environment Research (BER) – the real product of this program was NMR spectroscopy, which tried to image the behavior of atomic nuclei in a magnetic field and made scientists able to study the properties and the structures of the molecules. Magnetic resonance imaging (MRI) was not initiated till the 1971 – when Paul Christian Lauterbur encoded the spatial information into a nuclear magnetic resonance (NMR) signal using magnetic field gradients, the novel report was published in 1973 – while in 1977, the first MRI body scan of a human was performed by Raymond Damadian, Larry Minkoff and Michael Goldsmith. A lot of scientists contributed to the development of the MRI scanners – the ones accounted for that were Lauterbur and the researchers at Thorn-EMI laboratories,

England and Nottingham University, England. MRI became widely available in clinical practice in United States of America in 1984 – approved by Food and Drug Administration (FDA). Another product of DOE's nuclear medicine program was positron emission tomography (PET), which generates images by detecting energy given off by decaying active isotopes [1,3,4].

The techniques mentioned before (US, MRI, CT, PET, plus planar X-ray imaging) sum up the total of modern medical imaging techniques used in clinical practice. Each one produces images, which should not be limited to a fast 'diagnosis' or 'evaluation' from a physician, but contain a lot of information. This information can be analyzed with the help of sophisticated software and computers.
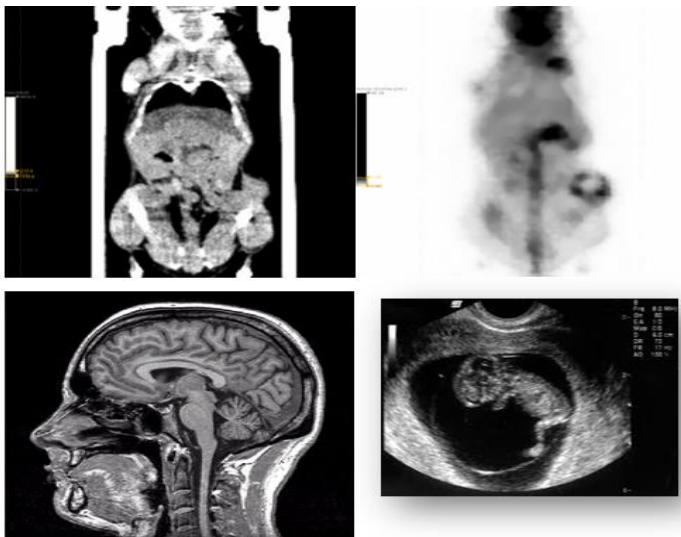
Figure 2. Images acquired with PET, CT, MRI [54] and US [55]. The upper left image is a CT image of a mouse, the upper right image is a PET image of a mouse, the lower left image is a MR image of a human and the lower right image is a US image of an embryo.

1.3 Positron Emission Tomography (PET)

1.3.1   Basic Principles of Positron Emission Tomography

As mentioned before, PET is a technique that generates images based on the radiation emitted from the decaying isotope (tracer) injected within the human body. These images are graphical representations of the distribution of the injected activity within the body. PET may also be used to assess the effectiveness of a treatment or a therapeutic scheme, to monitor the progress of a disease, or to track minor changes caused by a specific stimulus. Most of the time, the tracer (commonly tagged with oxygen-15, fluorine-18, carbon-11, or nitrogen-13) is administrated intravenously, through a peripheral vein. The total radiation dose in which the patient is exposed is similar to CT's dose. The typical time for a PET scan is approximately 10 to 40 minutes, while the patient remains fully clothed [5].

A well-known clinical application of PET imaging is the rate assessment of the glucose uptake in different parts of the body – 18-fluorodeoxyglucose (18F-FDG) a widely-used and well-tested radiolabeled analogue of the glucose is employed. 18F-FDG is frequently used to distinguish malignant from benign tumors (malignant tumors tend to metabolize glucose at higher rates than benign tumors), as well as for cancer staging. Other clinical applications of PET scan include: blood flow measurement, oxygen consumption in parts of the brain (e.g. checking the damage caused by a stroke or dementia), molecule tracking (e.g. chemical neurotransmitters, such as dopamine in Parkinson's disease) and cardiological applications (e.g. assessment of viable myocardium, before a transplant). Despite, the many different applications that PET has, it is used as a complementary diagnostic technique in the clinical routine – PET foundlings rarely can stand on their own [5].

Figure 3. Sibley Memorial Hospital's PET scanner [62].

1.3.2 Radiopharmaceuticals Used in PET Imaging

Besides 18F-FDG and fluorine-18, more well-studied and approved contrast agents and emitting radioisotopes (oxygen-15, nitrogen-13, carbon-11), which can be considered as biological tracers, do exist. Biological tracers are characterized by the ability to easily be substituted directly into the biomolecule without alternating its properties or function. Currently, there are only four FDA approved PET radiopharmaceuticals which are: 18-sodium fluoride (18F-NaF – used in bone imaging), 82-rubidium chloride (82Rb-RbCl – used in the assessment of the regional myocardial infusion, as well as in the diagnosis and localization of the myocardial infarction), 18-fluorodeoxyglucose (18F-FDG – used in the localization of tissues with abnormal glucose metabolic rates such as primary or metastatic malignant tumors), and 13-ammonia ($13N-NH_3$ – used in the monitoring and evaluation of the myocardial blood flow) [56].
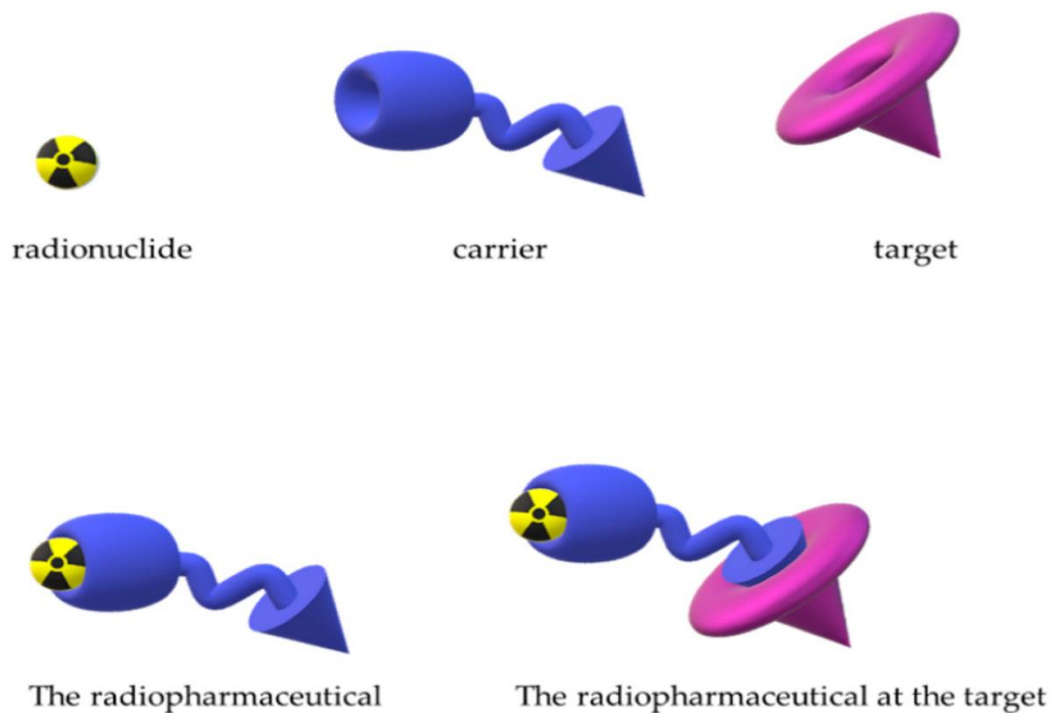
Figure 4. Display of the concept of design of a radiopharmaceutical drug with targeting properties [64].

1.3.3 18-fluorodeoxyglucose Positron Emission Tomography in Oncology

 The value of 18F-FDG PET in cancer diagnosis, staging and re-staging, monitoring, and treatment, is already well-proven. Optimal use of 18F-FDG include: early employment of the radiopharmaceutical from the very first stages of the diseases, continuous monitoring of the treatment progress for the accurate evaluation of the therapeutic scheme, and follow-up monitoring after the end of the treatment. Similar to any other medical imaging technique, limitations are evident.
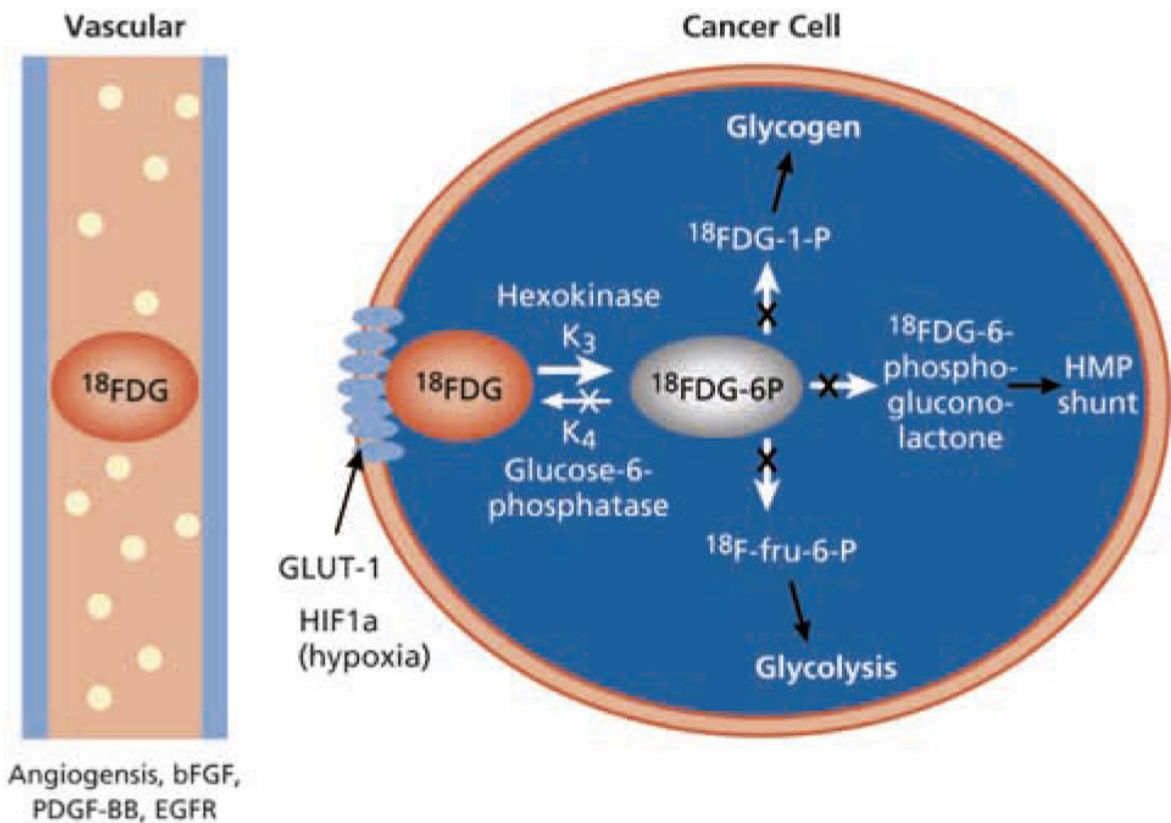
Figure 5. Display of the metabolism of 18F-FDG [65].

1.3.3.1 Benefits & Limitations of 18F-FDG PET

Highlighting the limitations, 18F-FDG tracer is highly absorbed by the cerebral grey matter of the brain and the heart in non-fasting patients, while a minor uptake of the tracer by the spleen, liver, and colon is observed. In addition, the excretion of the tracer happens through the urinary tract (kidneys, ureters and bladder). Hence, it is not optimal for the evaluation and monitoring of malignancies and metastases of tumors in the brain, heart and urinary tract. Another issue usually encountered with 18F-FDG PET is when applied to patients with high serum glucose level, the tumor to background (T:B) ratio may not be the most favorable because of the inevitable competition between the glucose analogue (18F-FDG tracer) and normal circulating glucose. As stated before, the sensitivity of this metabolic technique is highly dependent on the tracer's accumulation at the tumor site, while sensitivity is independent of tumor's structural characteristics. However, recent research data supports that histological types and subtypes – of both

primary tumor's and metastases' – in fact, they do influence the glucose uptake. Lastly, tumor size is another big issue, while all tumors with a size over 1 cm are always clearly seen, lesions with smaller sizes (size threshold strongly depends on the features of the scanner – huge intra-scanner variability) might be undetectable – lesions with a size ranging from 0.5 to 1 cm, their detection is not consistent [6].

On the other hand, tumors avidly absorbing glucose and tumors located into lungs, pleura, mediastinum, liver, spleen, skeleton, peritoneum, and digestive system, are easily detected because of the improved T:B ratio. 18F-FDG uptake is closely related to the tumor's viability and aggressiveness, while due to its absorption in tissues with high metabolic rate, 18F-FDG enables the measurement of tumor differentiation, heterogeneity (both intra- and inter-tumoral heterogeneity), and grade – high-grade and homogenous tumors (tumors usually expressing an inherent metastatic tendency) are characterized by higher 18F-FDG uptake rates, thus, 18F-FDG is considered as a potential biomarker of them – which means that they may come with a measurable degree of aggressiveness. While, it is already well-known that this metabolic technique's ability to detect lesions with smaller sizes (>5 mm) is questionable, it seems that under specific conditions (when sentinel node biopsy (SNB) is performed optimally) 18-FDG PET can detect microscopic nodal metastases with high accuracy, even at a very early stage [6].

1.3.3.2 Recent Advances of 18F-FDG PET in Different Cancer Types

1.3.3.2.1   18F-FDG PET in Head & Neck Cancer

Head and neck cancer are some of the most frequent cancer types met in the general population and – according to Parkin's et al. survey [8] – the 5th leading cancer by incidence. In the most challenging cases – unresectable tumors and/or locally advanced head and neck tumors – a combination of both radiotherapy (RDT) and chemotherapy (CHT) is a one-way road. Patients ending up in the last category usually have a poor prognosis, while in nearly 40% of the cases local recurrences are observed. Tumors

located in the oral cavity, pharyngeal or laryngeal sphere, are hard to be analyzed with 18F-FDG PET, because of their limited spatial resolution – while 18F-FDG PET cannot be used for the staging of the disease in the early stages (stages 0, I, II), also PET is unable to accurately define the tumor's involvement in adjacent structures (e.g. vessels, nerves, fat, muscle) [6]. Even though in the recent past PET imaging had a 'palliative' role – was mainly used in stage IV head and neck cancer patients to determine whether surgery or an aggressive therapeutic scheme was making sense – it now allows: quantification of the tumor's metabolic activity, better sensitivity and specificity in primary tumor's diagnosis, higher accuracy in the nodal staging of regionally advanced cancers, and better management of the therapeutic scheme (the treatment which a patient receives). Castelli et al. [7] reported in a 2016 review – using data related to the intensity of the tumor's metabolism, which is calculated from the Standard Uptake Volume (SUV) – that Metabolic Tumor Volume (MTV) was highly correlated with both overall survival and disease-free survival in head and neck cancer patients, while MTV's predictive value was superior in comparison to SUV's value [6,7].

1.3.3.2.2   18F-FDG PET in Esophageal Cancer

Esophageal carcinoma is one the most common (8th most common cancer worldwide in 2018) and lethal (6th leading cause of cancer-related death), with more than half of all the cases reported in China [10]. There are two main sub-types of the esophageal carcinoma: the esophageal squamous-cell carcinoma (ESCC) – more common in the developing world (countries with low human development index, e.g. low industrial development), also characterized of high metabolic rates (therefore avidly accumulates 18F-FDG) – and the esophageal adenocarcinoma (EAC) – more common in the developed world (countries with high human development index), usually characterized of high metabolic rates as well (except some types of EAC located at the caudal part of the digestive system). 18F-FDG PET imaging is incorporated in this type of cancer when an unsuspected distant metastasis is spotted – stage IV of the disease. However, acquiring valuable information about local lymph-node involvement through PET is rarely possible, because of the

localization of the node (usually extremely close to the primary tumor) – lower sensitivity than endoscopy US, but higher specificity. A combination of RDT and CHT is the most common approach against the esophageal locally advanced tumors. The real challenge doctors encounter when dealing with an ESCC/EAC patient is defining the radiation dose – as a lot of failures have been reported, making the control and the escalation of the dose, a difficult case. Yu et al. [9] performed dose escalation in esophageal cancer patients – cancer had been evaluated and displayed using 18F-FDG PET – up to 70 Gy using simultaneous integrated boost (SIB) technique, without causing any life-threatening issue – acute toxicities were noticed and dealt with success, however not enough information was gathered about late toxicities and long-term issues [6,9].

1.3.3.2.3   18F-FDG PET in Lung Cancer

 In 2018, lung cancer was the most common and lethal cancer – being responsible for the 17% of the newly reported cancer cases, and nearly 25% of all cancer-related deaths – according to Bray et al. [12] estimation using GLOBOCAN. The most usual lung cancer type met in patients is non-small cell lung cancer (NSCLC), with adenocarcinoma being the major histological subtype of NSCLC – while further subtypes of lung adenocarcinoma have been identified, which can be split into different grades based on prognosis. 18F-FDG PET can be used for diagnosis (even though CT is usually preferred, because of the limited spatial resolution of PET – small lesions, under 1 cm will be missed), staging (PET allows the better and more accurate definition of the tumor, while it is optimal for differentiating N0-N1 stages from N2-N3 stages – which are critical in the determination of patient's management – and N stages from M stages), and monitoring of cancer's progress. One of the most important applications of PET imaging in lung cancer is the accurate and precise targeting of the tumor's total volume (meanwhile exposing healthy tissue to minor radiation doses in contradiction with the CT) in high-risk patients (mostly due to impaired lung function, atelectasis, and/or pleural effusion), whose capability to radiotherapy is questionable. Another famous and recent application of PET imaging is the utilization of PET's standardized uptake value (SUVmax) feature in the prognosis of

NSCLC – it is already found, there is a strong correlation between the SUVmax and the pathological subtype and the histological grade of lung adenocarcinoma. Furthermore, Nakamura et al. [13] reported a strong correlation between the pre-operative SUVmax and the histological adenocarcinoma subtype and its subgroups, as well. While, Sun et al. [11] reconfirmed the previous SUVmax and lung adenocarcinoma grade relationship, and proposed the incorporation of 18F-FDG PET as a non-invasive diagnostic tool, which can also contribute to the patient's pre-operative treatment management [6,11].

1.3.3.2.4   18F-FDG PET in Breast Cancer

Breast cancer was the most frequently diagnosed cancer amongst women all over the world, while it achieved the highest mortality rates in 11 regions of the world – according to Ferlay's et al. [15] survey, using GLOBOCAN (Global Cancer Observatory). Even though they already exist extremely specialized techniques for breast cancer T-staging and screening (e.g. ultrasonography and mammography), 18F-FDG PET still offers some advantages. In pretreatment staging, it may provide: information about the metabolic activity of the tumor (lobular breast cancer are characterized of higher 18F-FDG uptake than ductal carcinomas), additional information to supplement an inconclusive MRI (even though dynamic contrast-enhanced MRI is more sensitive than PET in the spotting of small lesions – under 1 cm), and useful information to guide a biopsy in fatty breasts and/or fibrotic tissue initiated from a surgery/radiation therapy. 18F-FDG PET is also useful in the detection of distant metastases (stage IV of the disease), in cases of loco-regionally advanced disease (e.g. SNB positive), and for patients that are good candidates for neoadjuvant CHT or RDT-CHT. Neoadjuvant chemotherapy (NAC) promotes a pathological complete response (absence of invasive cells in the primary tumor and the lymph nodes), which has been linked with higher survival rates – NAC seems to offer better results in the most aggressive types of breast cancer (e.g. triple-negative), while it has been noticed that the pre-treatment assessment of various clinical and biological breast cancer characteristics can lead to a better prognosis. 18F-FDG's PET SUVmax feature – high SUVmax values have already been linked with lower survival rates and high

histological grade – seems to have a prognostic value in breast cancers (more limited in ER+/HER2- than in other subtypes – luminal tumors are characterized of lower 18F-FDG uptake). In a 2017 clinical study, Groheux et al. [14] reported that no additional information was gained through textural analysis of the 18F-FDG PET image, while the conclusions (PR-negative or lobular invasive breast cancer patients have a poorer prognosis than PR-positive or ductal carcinoma breast cancer patients) were made based on histopathological characteristics. Another less-popular feature offered by 18F-FDG PET is the detection of brown adipose tissue activation – brown adipose tissue activation has been proposed (lately) to have a possible major role in the development of breast cancer. In fact, Pace et al. [16] recently identified a strong correlation between brown adipose tissue activation and positive-known prognostic factor in breast cancer (e.g. the relationship between intermediate tumor grade and luminal B cancer type) [6,14].

1.3.4 Recent Advances of 18F-FDG PET in Infectious & Inflammatory Diseases

Applications of 18F-FDG PET/CT in oncology (as mentioned before) are already well-known for over a decade. However, this hybrid metabolic imaging technique has a tone of potential (and in some cases current) applications in the evaluation of a huge range of pathological conditions, which are related to infection and aseptic inflammation. The biological mechanism which FDG is delivered to the tissues/cells of interest has been studied extensively – FDG is an abnormal (not met in nature) analogue of the glucose, which enters the cell the same way as glucose (to join in the glycolytic metabolism) – ending up accumulating in the tissues with higher metabolic rates. Hence, it was expected that 18F-FDG PET/CT would have more clinical applications – beyond oncological imaging and patient's management – in different pathological fields, including suspected infection and inflammation, central nervous system disorders, and cardiovascular diseases [17].

1.3.4.1 Molecular Basis of 18F-FDG PET Imaging in Infection & Inflammation

As stated before, 18F-FDG is introduced to the cell and its metabolic pathways with physiological processes similar to glucose – 18F-FDG's uptake takes place into the cellular surface by special glucose transporters (especially glucose transporter-1 – GLUT1), while it is metabolized to 18F-FDG-6-phosphate from hexokinase (a rate-limiting enzyme of glycolysis). Heinrich Warburg [18] was the first who noticed that cells of the lesions were characterized by a more intense metabolism than healthy cells (through a non-oxidative breakdown of glucose – the "Warburg" effect). Many years later, Kubota et al. [19] reported that the inflammatory cells (e.g. macrophages), located at the peripheral sites of the tumor, can take on larger quantities of 18F-FDG than the tumor's cells. In general, there are 3 main inflammation phases: early vascular phase (characterized by tissue hyperaemia, enhanced vascular permeability, and inflammatory mediator release – this increased tissue perfusion, to the affected sites, results in greater 18F-FDG uptakes), acute cellular phase (characterized by active cell recruitment, migration, and proliferation at the affected area – burst release of a huge variety of cytokines, upregulation of GLUT-1 and GLUT-3, increase in hexokinase activity and GLUTs amplification to their substrates are enhancing the pathways of glycolysis, thus resulting in greater FDG uptake), and late cellular/healing phase. Multiple mechanistic similarities between inflammatory and malignant can be spotted, which end up in the same result – greater 18F-FDG uptake at the affected sites [17].
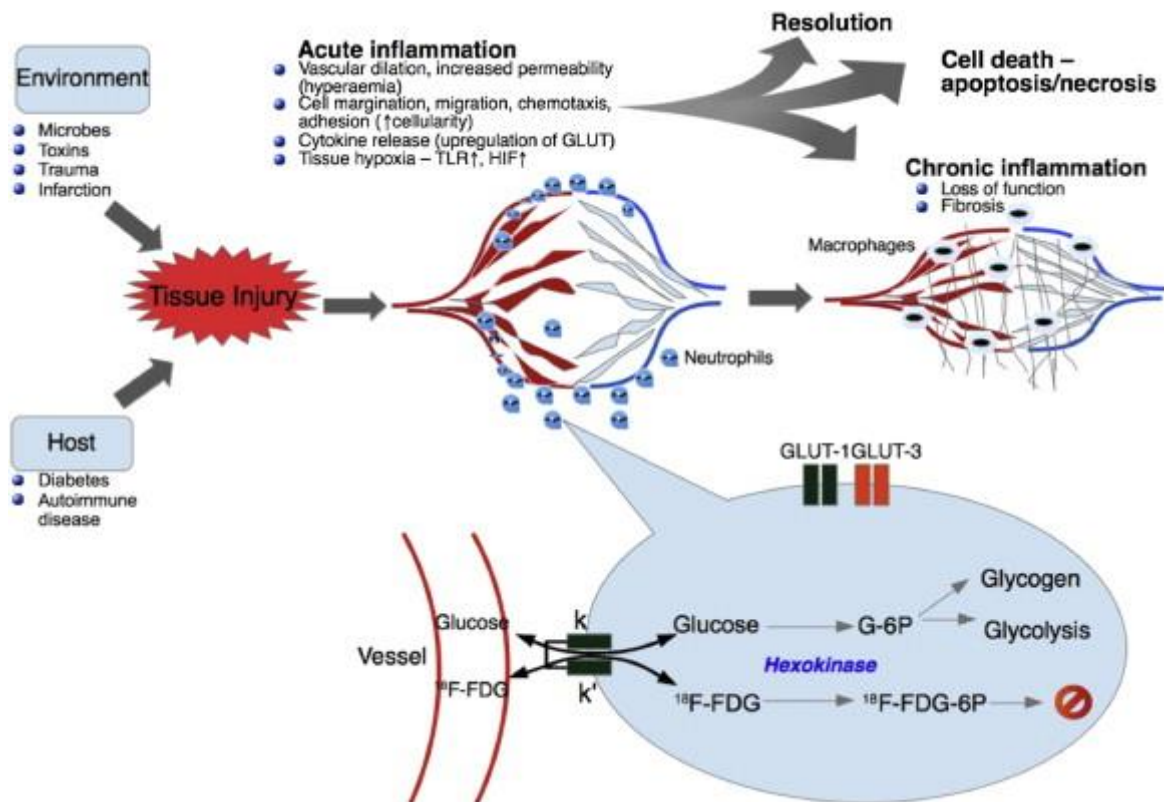
Figure 6. Display of the introduction and metabolism of the 18F-FDG into the inflammatory cells – inflammation caused by tissue injury [17].

1.3.4.2 18F-FDG PET in Fever of Unknown Origin (FUO)

FUO was firstly described by Petersdorf et al. [20] as a condition in which patient experiences temperatures greater than 38.3oC on several occasions, that lasts for at least 3 weeks, without an obvious cause after a week of patient's hospitalization (currently changed to either 3 or more outpatient visits or 3 days of inpatient investigations) – it remains still today a diagnostical enigma, which can be extremely dangerous for high-risk population (e.g. the elderly, the young, and the immunocompromised). FUO has 4 different clinical phenotypes: classical FUO, nosocomial FUO, immunocompromised FUO, and (Human Immunodeficiency Virus) HIV-related FUO. Even though there are a lot of different biological factors that can lead to FUO, there are 3 main groups that "stand out": infection, malignancy, and inflammatory disorders – in 50% of the cases, no distinct cause was defined. 18F-FDG PET has been clinically useful in a lot of FUO cases, while it is

characterized by extremely high sensitivity, specificity, accuracy, and negative predictive value (in some cases up to 100%). Furthermore, 18F-FDG PET can be integrated with CT and/or MRI, to supplement these morphological techniques with information about the metabolic background of the affected tissue – meanwhile, PET imaging offers: whole-body coverage, higher sensitivity, hybrid imaging, safety to patients with renal failure or allergy to iodinated contrast media, and may guide a biopsy or used for treatment planning and/or monitoring, as well. In a 2013 clinical study performed by Kouijzer et al. [21], 18F-FDG PET was compared to the conventional radionuclide test used in FUO cases (radiolabeled white cell scintigraphy and gallium-67 citrate scintigraphy – imaging techniques suffering from low spatial resolution, poor target-to-background contrast, and highly time-consuming, while they require a lot of staff to involve), proving PET's diagnostic superiority [17].
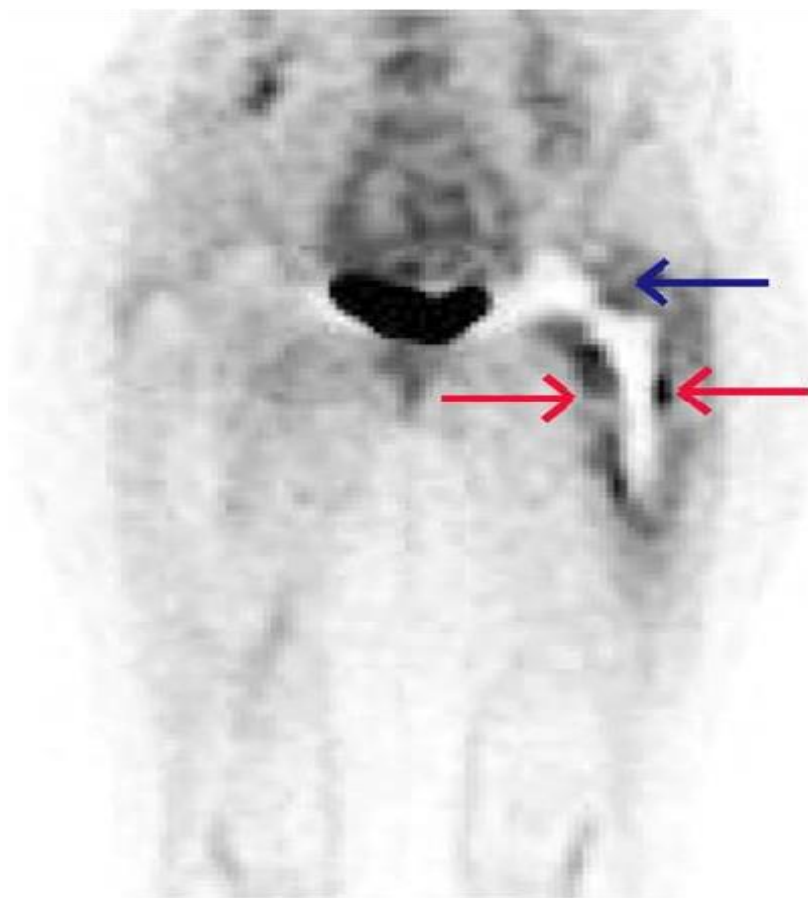
Figure 7. Display of the inflammation in patient with prosthetic hip joint infection [67].

1.3.4.3 18F-FDG PET in Joint Prosthesis Infection

Arthroplasty surgery has dramatically improved the quality of life of modern people (especially of elderly and severely injured people), by incorporating hip and knee prostheses and replacing/fixing the damaged joints – restoring normal joint function (in some cases over 90-95% function restoration). Nearly 25% of these surgeries will be problematic (mechanical loosening issues), and the patients would have to repeat the surgeries. An implication rarely seen in this kind of surgery is the prosthetic infection (occurring in less than 1% of patients having primary knee or hip arthroplasty) – prosthetic infection can lead to the septic or aseptic loosening, defining which one occurs can be decisive of patient's prognosis and the revision surgery. Traditionally, leukocyte scintigraphy integrated with marrow scintigraphy is being used for the diagnosis of prosthetic infections, showing some incredibly high sensitivity, specificity, and accuracy. On the other hand, 18F-FDG PET (even though it can be used for the diagnosis of "conventional" infectious and inflammatory diseases with excessive superiority over other scintigraphy techniques) seems to be pretty inconsistent with its performance in the diagnosis of prosthetic infections – sensitivity ranging from 28% to 91% and specificity ranging from 9% to 97%, according to van der Bruggen's et al. [22] survey. In fact, Aksoy et al. [23] conducted clinical research reconfirming that 18F-FDG PET/CT is not an infection-specific technique, however FDG-labelled leukocyte 18F-FDG PET/CT may have some potential to outperform the conventionally labelled leukocyte scintigraphy in infection imaging [17].

1.3.4.4 18F-FDG PET in HIV-Related Infections

In a 2018 report from UNAIDS [24], it was estimated that 37.9 million people worldwide were found HIV-positive, tens of millions of people had died of HIV-related causes, while nearly 2.5 million people would be diagnosed HIV-positive every year. It is safe to claim

that HIV is a modern epidemic. In a 2013 clinical study, Sathekge et al. [25] reported that there is a strong correlation between patient's lymph nodes capability to take on 18F-FDG (a "march of activity" is usually observed in these nodal groups) and viral replication – and an inverse correlation with CD4+ counts. As mentioned before, 18F-FDG PET/CT performs well in the localization of the source of infection in HIV-related FUO. One of the greatest applications of 18-FDG PET/CT in HIV-positive patients is in the determination of whether the patient suffers from a cerebral lymphoma – in this 18F-FDG PET/CT is usually integrated with MRI – or opportunistic infections (e.g. toxoplasmosis – characterized by increased intra-tumoral FDG uptake) [17].
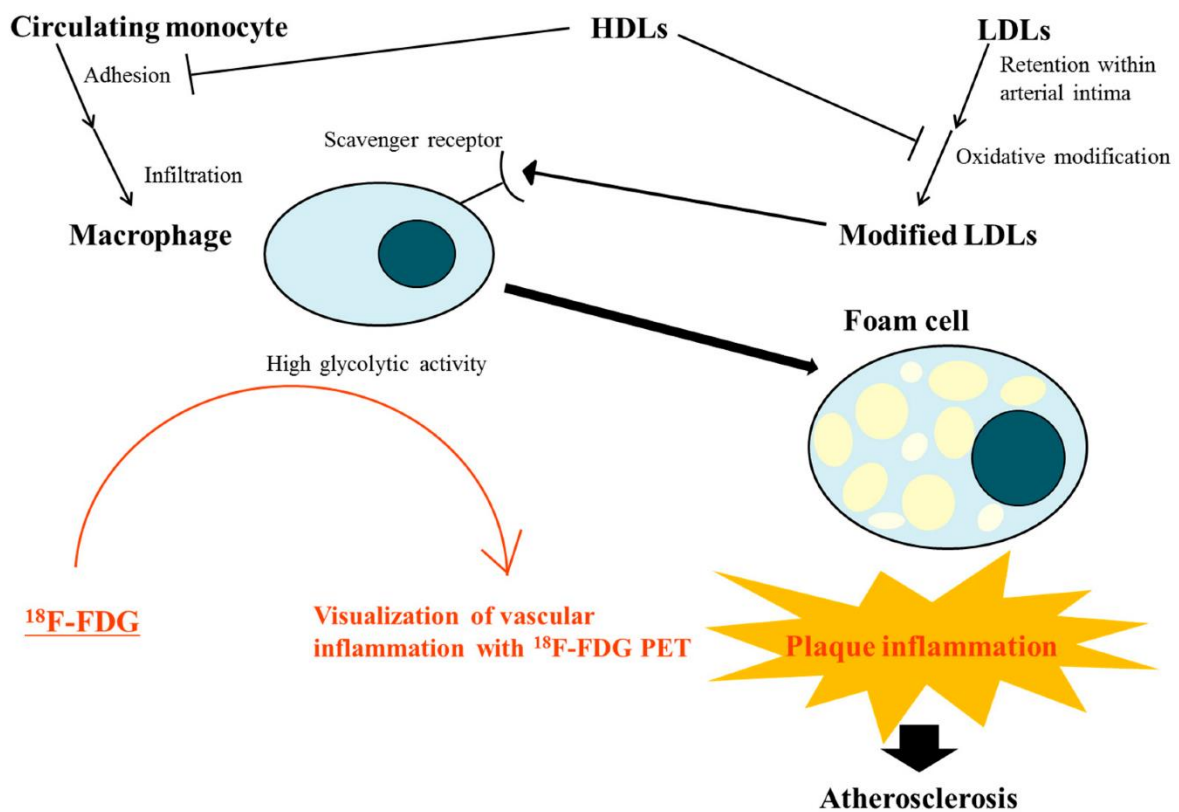
Figure 8. Display of the way of visualization of the arterial inflammation with the employment of 18F-FDG [66].

1.3.4.5 18F-FDG PET in Arterial Inflammation

Atherosclerosis is a systemic, inflammatory disease of the arterial wall which avidly accumulates FDG, therefore it can be quantified with 18-FDG PET. The inflammatory activity on the arterial wall – products of this activity are the arteriosclerotic lesions (when in advanced phase, arteriosclerotic lesions are calcified in the point they can be easily spotted with CT) – has already been associated with both fatal and non-fatal cardiovascular events. Vascular calcification is linked with high cardiovascular morbidity and mortality, both in a high-risk population (e.g. patients with chronic kidney diseases/diabetes) and the general population (healthy people). This calcification and inflammation of the arterial can be measured and quantified with 18F-FDG PET/CT (because of the enhanced metabolic pathways on the affected sites) – an important application especially for lung cancer patients who are characterized from the same risk factors as arteriosclerosis' patients (e.g. age, smoking, diet). In a 2020 study conducted by Garcia et al. [26] – 18F-FDG PET/CT images of lung cancer patients were evaluated through texture analysis – it was noticed that, although patients with bronchogenic carcinoma had a high degree of arterial wall inflammation, it had no impact in the generation of cardiovascular events, while Calcium score did affect them [17,26].

1.3.5 Novel PET Imaging Tracers/Radiopharmaceuticals

Recent advances in Pharmaceutical Technology and Nanotechnology seem promising, as a new era of tools – ranging from contrast agents to tracers – is under development and is going to enable the rise of new fields in medicine (one of those fields is the field of Radiomics). Some recent novel PET imaging tracers are quoted below which applications are ranging from imaging and tracking of Alzheimer's disease and evaluation of myocardiopathy to the assessment of tryptophan metabolism. None of them has been approved by the FDA yet – still under research and excessive testing.

1.3.5.1 18F-F-Polyglucose Nanoparticles – Macroflor

 In vivo tracking of macrophages had been an area of interest for quite a long time. Macrophages are a type of white blood cells of the immune system which have a major role in both the normal cellular function (e.g. digestion of cellular proteins or residues) and the defense of the body against pathogens and cancer (e.g. digestion of pathogens/cancer cells). They are located in all the tissues of the body – their counts may increase when tissue is under stress or disease. Vast numbers of inflammatory macrophages are directly related to a wide variety of serious conditions (e.g. atherosclerosis, myocardial infarction, heart failure), as they can accumulate and damage the tissue. Keliher et al. [27] incorporated a class of modified polyglucose nanoparticles known as macrins (polymers of lysine cross-linked low molecular weight carboxymethyl polyglucose, consisting of 22 glucose units each) with a mean size of 5 nm, labeled with 18-Fluorine – the emerged nanoparticle was named Macroflor. Macroflor was designed for the targeting and tracking of the macrophages of the cardiovascular system and effective renal clearance. Both of these claims were confirmed in three animal models. Also, it provided accurate, quantitative and specific PET data on inflammation in atherosclerotic plague and ischemic myocardium – surpassing the problems of previously developed nanoparticles with high affinity for macrophages which were reported to have limited target-to-background ratios, because of their long circulation times and high blood pool activity [27].

1.3.5.2 ^{68}Ga-Galmydar

 Another unconventional PET radiotracer targeting organs of the cardiovascular system is ^{68}Ga-Galmydar. ^{68}Ga-Galmydar was designed to assess whether a patient treated with Doxorubicin (DOX) is in danger of DOX-related heart failure. DOX is a common and widely available chemotherapeutic drug used for treating in numerous cancer cases (coming in various pharmaceutical forms and technologies – including nano-form, encapsulated into liposomes e.g. DOXIL®). However, DOX has been associated with acute and chronic

cardiac anomalies (e.g. aberrant arrhythmias, ventricular dysfunction), even heart failure – because of its indiscreet antimitotic and cytotoxic effect (inhibition of DNA transcription, abnormal protein folding, ROS generation, oxidative stress) [29]. Sivapackiam et al. [28] used micro-PET imaging 5 days post-treatment in rats to evaluate the ability of ^{68}Ga-Galmydar to notice cardiomyopathies related to DOX, as well as post-imaging quantitative biodistribution studies. It was reported that ^{68}Ga-Galmydar could indeed provide valuable information about DOX-linked and reversible metabolic changes, even at earliest stages – surpassing the problems of the most usual medical technique used in relative cases, echocardiography (even though it is cost-efficient and widely available, echocardiography is an invasive technique which cannot identify the initiation of DOX-related heart failure at early stages) [28].

1.3.5.3 ^{18}F-PI-2620

^{18}F-PI-2620 is a novel tracer used in PET imaging, with unique properties – it can accumulate and bind selectively and specifically to tau aggregates (more specific it binds to both 3-repeat and 4-repeat tau isomorphs). Tau is a protein which aggregates and agglomerates have been linked with major neurodegenerative diseases, including Alzheimer's disease (AD) – which is the leading cause of dementia and one of the biggest health concerns globally, according to a survey conducted by World Health Organization (WHO) [31]. AD is a slowly progressive disease (in most cases, it takes up to 10 years or more for the first symptoms to appear) and is characterized by the presence and continuous growth of β-amyloid plaques in the extracellular space and hyperphosphorylated tau neurofibrillary tangles inside the nerve cells. Definite diagnosis of AD is not a simple case, as it comes with a lot of changes and consequences in the daily life of the patient. However, its immediate diagnosis at early stages is crucial for the monitoring and development of the disease. Tau deposition in β-amyloid-positive patients paired with clinical evaluation is the most "sure bet", currently available in the diagnosis of AD. Besides that, tau tracking and imaging may also provide useful information about tau's physiopathology (other tau-accumulation-related diseases:

progressive supranuclear palsy, corticobasal degeneration, Pick disease) and its effect in cognitive function. Mueller et al. [30] tested ^{18}F-PI-2620 effectiveness in a clinical trial (featuring both AD patients and healthy individuals), reporting ^{18}F-PI-2620's superiority in comparison with other tau PET imaging tracers (as it provides high-quality images and has an excellent signal-to-noise ratio in AD patients – first-generation tracers were characterized by off-target binding in several brain regions, including basal ganglia and choroid plexus). Furthermore, they were able to correlate tau accumulation with the overall neurocognitive performance [30].

1.3.5.4 6-^{18}F-α-methyl-L-tryptophan

 6-^{18}F-α-methyl-L-tryptophan is another novel tracer used in PET imaging. In fact, 6-^{18}F-α-methyl-L-tryptophan is a "reworked" and enhanced version of an older PET tracer, α-[^{11}C]Methyl-L-tryptophan which has been used extensively in the assessment of tryptophan metabolism in cancer, epilepsy, migraine, tuberous sclerosis, and autism, as well as in various psychiatric illnesses (e.g. schizophrenia, depression). Initial application of the α-[^{11}C]Methyl-L-tryptophan was the quantification and measurement of the metabolism of tryptophan into serotonin inside the brain. However, this tracer had 3 major limitations: the short half-life of the radionuclide (^{11}C) used for the labeling which is approximately 20.4 mins (thus not allowing the more complex synthesis and transportation to PET centers where an onsite cyclotron is not available), irreversible uptake of the radio-labeled tracer inside the brain, and its extensive metabolism along the kynurenine pathway – it is already well-known that up to 95% of tryptophan undergoes oxidative metabolism along the kynurenine pathway, therefore α-[^{11}C]Methyl-L-tryptophan is more likely to reflect the activity of the kynurenine pathway in a diseased state where the pathway has been induced. The kynurenine pathway has been an area of interest for quite a long time now, as inhibition of the pathway (inhibitors targeting the main enzyme of the pathway, indoleamine 2,3-dioxygenase(IDO) – α-[^{11}C]Methyl-L-tryptophan is a substrate of IDO, thus showing increased uptake rate in lung, breast, and brain tumors) via immunotherapy has been suggested as a putative therapeutic

approach. However, induction of this pathway may lead to increased accumulation of kynurenic acid and production of quinolinic acid inside the brain, a condition that has been linked with suicidal behavior and depression. Krasikova et al. [32] conducted a series of in vivo, in vitro, and in silico experiments using 6-^{18}F-α-methyl-L-tryptophan, reporting tracer's improved stability (robust to in vivo defluorination), safety (same levels of brain exposure with α-[^{11}C]Methyl-L-tryptophan) and kinetic profile (same as α-[^{11}C]Methyl-L-tryptophan's kinetic profile). Also, the research group suggested that 6-^{18}F-α-methyl-L-tryptophan uptake into the brain occurs through the LAT1 transporter rather than kynurenine or serotonin metabolic pathways, as 6-^{18}F-α-methyl-L-tryptophan is not a substrate of IDO [32].

1.3.5.5 ^{68}Ga-Labeled Prostate-Specific Membrane Antigen

^{68}Ga-labeled prostate-specific membrane antigen (PSMA) is a novel tracer – recently introduced – used in PET imaging of prostate cancer in patients diagnosed at high-risk and patients whose treatment failed (mainly because of biochemical failure). However, it seems that ^{68}Ga-PSMA can accumulate in a huge variety of solid tumors (including hepatocellular carcinoma (HCC)). Hence, this may reflect ^{68}Ga-PSMA's future potential for HCC imaging and monitoring. HCC was one of the most common and lethal cancers worldwide (5th most common cancer and 3rd most common cause of cancer-related mortality) in 2013, according to a survey conducted by Mittal et al. [34]. Amongst the highest-impact risk factors for HCC, cirrhosis due to viral hepatitis or nonalcoholic fatty liver can be met, while in more than 30% of the cases remote metastases have been reported. Diagnosis of the HCC at early stages is crucial and can lead to a better prognosis. Imaging and monitoring of the disease are both essential and challenging, as relapse of the disease can be easily confused with coagulative necrosis, hematomas, abscesses, and fluid or bile collections. 18F-FDG PET imaging applications are limited in this type of cancer, as fewer than half of the cases are reported to be 18F-FDG-avid tumors. Kesler et al. [33] carried out a prospective pilot study evaluating the ^{68}Ga-PSMA's potential as an imaging agent of HCC, reporting noticeably increased rates of uptake of

^{68}Ga-PSMA inside HCC lesions (as most of them are hyper-vascular and ^{68}Ga-PSMA can be easily trapped inside tumor's micro-vessels). Hence, ^{68}Ga-PSMA's imaging was superior in comparison to 18F-FDG's imaging [33].

1.4 Texture Analysis

Applications of texture analysis into medical imaging have gained a lot of attention, recently. During the last decade, an increased number of papers, using texture analysis of PET/CT images for the quantification of intratumor uptake heterogeneity, has been published. The applications are ranging from a more accurate prognosis or diagnosis to the linkage of certain texture indices with the tumor's biochemical characteristics. However, as this is a newly emerged field there are a lot of challenges to be addressed (e.g. need for protocol standardization, equal comparison of the research findings, bioethical issues, etc.) [35].

1.4.1   Image Classification: the Beginnings of Texture Analysis

The very first steps of texture analysis can be traced back to 1973 when Haralick et al. [36] tried to identify some reliable texture features with the ultimate goal to use them for image classification. There are three main types of features containing meaningful information for image classification:

- **Spectral features:** contain information about the tonal differences in the visible or infrared part of the electromagnetic spectrum
- **Textural features:** contain information about the spatial distribution of tonal variations within a band
- **Contextual features:** information coming from the surrounding regions of the area of interest.

By far, the most important characteristic is texture. The texture is an innate property of all surfaces which provides information about the structure of the surface, as well as the surface's relationship with the surrounding environment. Texture, tone, and context are always present in any image and are characterized by a "wave-particle relationship" – occasionally one can be dominant to the others. Even though the images that Haralick used were two dimensional (2D) arrays composed from 2D pixels, while the modern medical images used in texture analysis are three dimensional (3D) models composed from 3D voxels, the same concepts and principles are applied till today [36].
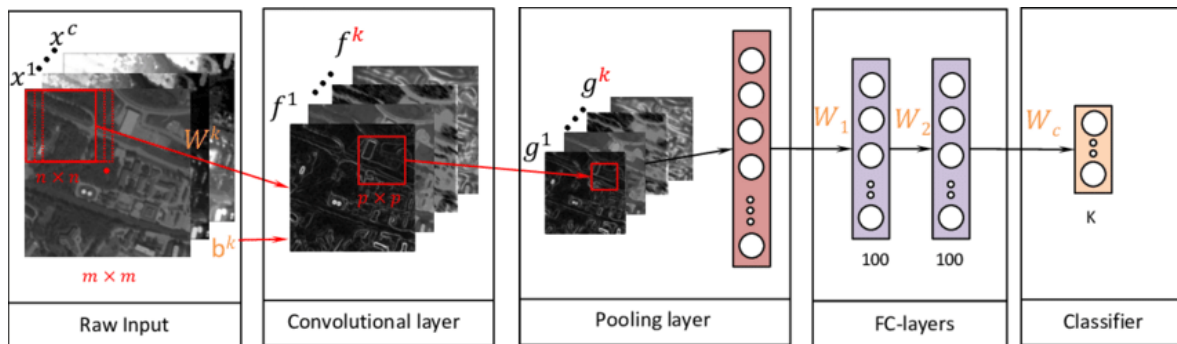
Figure 9. Overview of the process used for per-pixel image classification [68].

1.4.2 Texture Analysis Approaches

Texture analysis utilizes mathematical formulas which can provide information about the grey-level intensity (grey level indicates the brightness of a pixel/voxel) distribution and spatial organization in the image. In other words, texture features (indices) are mathematical parameters coming from pixels'/voxels' properties. There are several methods which can be incorporated to compute texture features, mainly divided into three major groups:

- **Statistical methods:** the texture is represented using well-defined primitives. These methods describe the grey-level intensity distribution as well as the relationship of grey-level values in the image. Features extracted with statistical methods are divided into three main categories: the First-Order features

(Histogram, Conventional and Shape indices), the Second-Order features (Grey-Level Co-occurrence Matrix (GLCM)), and the Higher-Order features (Grey-Level Run Length Matrix (GLRLM), Neighborhood Grey-Level Difference Matrix (NGLDM), Grey-Level Zone Length Matrix (GLZLM)) – the computation of the texture features in this investigation was done using Statistical methods

- **Model-based methods:** the texture is represented mostly using sophisticated and complex mathematical models – fractal or stochastic models, incorporating different algorithms

- **Transform-based methods:** the texture properties are analyzed in a different domain, such as the frequency or the scale domain – methods featuring in this class: Fourier transform, wavelet and Gabor filters, and the Laplacian transform of Gaussian filter. These methods allow the transformation of the whole image, making it possible to extract statistical features on a multiscale approach.

Texture features are always calculated for a given region of interest (ROI) which is set manually by the user and may be expressed as either a single value for the whole ROI or as a texture map of the ROI [37].
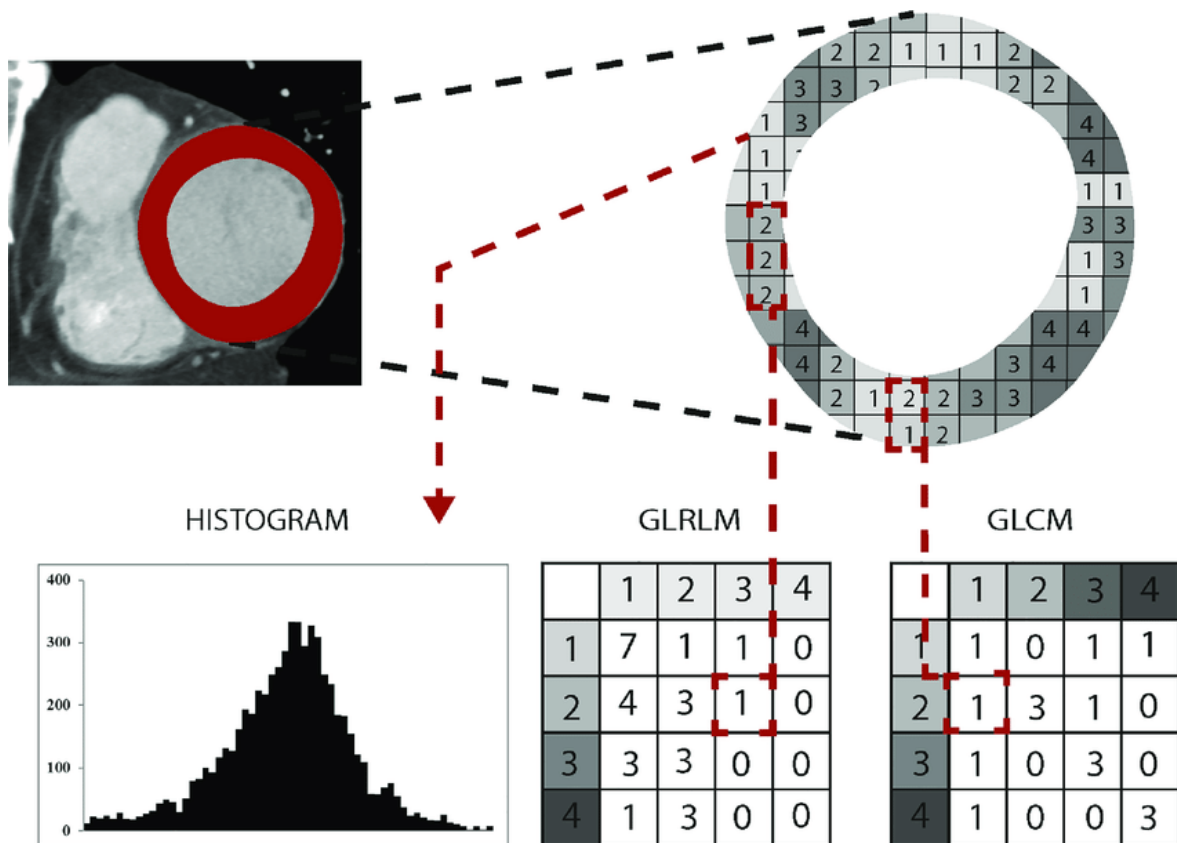
Figure 10. Schematic overview of the texture analysis process [69].

1.4.2.1 First-Order Features

Conventional, Shape, and Histogram features belong to this class.

- **Conventional indices:** provide information related to the ROI's FDG uptake ability. Features usually calculated in this category include: standard uptake value's (SUV) minimum, mean, and maximum value, and total lesion glycolysis (TLG) value – which is a PET-specific feature and is the product of SUV mean value by total volume (mL).

- **Shape indices:** provide information related to the ROI's shape properties. Features usually calculated in this category include: sphericity, compacity, and total volume (mL/voxels).

- **Histogram indices:** provide information about the global distribution of grey-level intensities in terms of spread, symmetry, flatness, uniformity, and randomness.

Features usually calculated in this category include: skewness (measures the asymmetry of the grey-level distribution), kurtosis (compares the shape of the grey-level distribution (peaked or flat) to a normal distribution), entropy (measures the randomness of the grey-level distribution), energy (measures the uniformity of the grey-level distribution), mean value, and variance [37,38].

1.4.2.2 Second-Order Features

Grey-Level Co-occurrence Matrix (GLCM) features belong to this class. GLCM – firstly introduced by Haralick et al. [36] – provide information about the spatial relationship between pixels/voxels in a certain direction (GLCM is run over 4 directions for 2D ROIs and 13 directions for 3D ROIs), describing the properties of uniformity, homogeneity, randomness, and linear dependency of the image. Features usually calculated in this category include:

- **Homogeneity:** measures the uniformity of the grey-level voxel pairs distribution in the ROI
- **Energy:** measures the textural uniformity or second angular moment of the image – the grey-level voxel pairs repetitions
- **Contrast:** measures the local variation of the grey-level intensities – shows the linear dependency of grey-level intensities of two neighboring voxels
- **Correlation:** measures how correlated the grey-level value of a voxel is to its neighboring voxel grey-level value
- **Entropy:** measures the randomness of the grey-level intensity distribution of the voxel pairs – provides information about the content of the image (e.g. high entropy matrix corresponds to an image, in which there are no grey-levels intensities preferred into the voxel pairs (for a given distance))
- **Dissimilarity:** measures the variation of the grey-level distribution of the voxel pairs [37,38].

1.4.2.3 Higher-Order Features

Grey-Level Run Length Matrix (GLRLM), Neighborhood Grey-Level Difference Matrix (NGLDM), and Grey-Level Zone Length Matrix (GLZLM) features belong into this class.

- **GLRLM:** provide information about texture in a certain direction (GLRLM is run over 4 directions for 2D ROIs and 13 directions for 3D ROIs), where more short runs are accounted to fine texture whilst more long runs are accounted to coarse texture with different grey-level values. Features usually calculated in this category include: short-run emphasis (measures the distribution of short homogenous runs), long-run emphasis (measures the distribution of long homogenous runs), grey-level emphasis (measures the distributions of runs composed of voxels with high/low grey-level values), grey-level and run length non-uniformity (measures the non-uniformity of the grey-level value or the length of the homogenous runs, respectively), and run percentage (measures the homogeneity of the homogenous runs) [37,38].

- **GLZLM:** provide information about the regional intensity variations or the distribution of the homogenous regions. Features usually calculated in this category include: short-zone emphasis, long-zone emphasis, grey-level emphasis, grey-level and zone length non-uniformity, and zone percentage. GLZLM indices are almost identical to GLRLM indices, with the difference that in the GLZLM's case the features concern zones instead of runs [37,38].

- **NGLDM:** provide information about the spatial relationship between three or more pixels/voxels. Features usually calculated in this category include: coarseness (measures the level of spatial rate of change in intensity of the grey-level of voxels), contrast (measures the intensity of grey-level difference between neighboring regions), busyness (measures the spatial frequency of changes in the intensity of grey-level of the voxels), complexity, and texture strength [37,38].

1.4.3   Factors Affecting the Computation of the Texture Features

The computation of the texture features is a vulnerable-to-errors process which can be easily altered by general variables, including factors related to the process itself and ROI identification (General factors), or factors specific to the image acquisition technique (Image-specific factors) [37].

1.4.3.1 General Factors

There is no acknowledged standardization of the texture features computation which can be affected by a huge variety of factors ranging from the SUV discretization criteria and resampling, the origin of the matrices and their properties to the texture analysis software package employed. Most of the features frequently show low-to-no robustness and low repeatability and reproducibility when even the slightest of a change occurs to one of these factors – it is rare for a texture feature to be characterized as stable and reliable, thus, it is extremely hard to propose a feature as a potential candidate for biomarker.

- **Lack of standardization:** since a "golden standard" for the computation of the texture features does not exist, this leads to a never-seen-before heterogeneity between studies – making it hard to compare the results and generate reliable conclusions.
- **Grey-level discretization:** the choice of the total number of bins, it is one of the most important factors when it comes to the computation of the texture features. In PET imaging a high-number-of-bins acquisition is preferred – recent research data suggests at least 32 discrete grey-level intensity values but no more than 64. Another issue is whether a fixed number of bins (FNB – the total SUV of the image is separated in a certain number of bins, then the texture features are calculated according to that separation, e.g. image's total

SUV is split into 64 bins, ROI's ΔSUV is split into 10 bins) or a fixed bin size (FBS – the ΔSUV of the ROI is separated in a certain number of bins, e.g. ROI's ΔSUV is split into 64 number of bins, image's total SUV is split into 200 bins) should be used. Even though current data suggests FBS should be preferred over FNB, a more in-depth investigation is required.

- **Isotropic resampling of the image:** this factor does not affect the computation of the First-Order features (since their computation is based on the distribution of the grey-level intensities into the histogram) but it does affect the Second-Order and Higher-Order features' computation (since the distance vector usually has a major role in the computation of these features). Isotropic resampling is suggested for 3D ROIs and the computation of Second-Order and Higher-Order features.

- **Non-standardized nomenclature:** there is no standard nomenclature for the texture features (e.g. same parameter in different studies may refer to different methodologies). Since Hatt et al. [35] addressed this issue, some improvements have been made.

- **Directionality in texture matrices:** computation of GLCM and GLRLM features is based on the directions in the matrices. The total number of directions, choice of directions and, whether or not the results are averaged (over the total number of directions) may have a major impact on the final result.

- **Specific parameters for texture matrix formulation:** in GLCM's case the vector distance has a crucial role in the computation of the features. Most of the time, a distance of 1 pixel/voxel is preferred to examine the relationship between neighboring pixels/voxels. However, in some cases, higher values may be preferred.

- **Use of already available packages:** there is a huge variety of available software, making texture analysis easily-accessible and user-friendly. However, different software uses different mathematical formulas to calculate the same parameters or may use different nomenclature for the same function which leads to a couple of issues – e.g. unfair or unreliable comparison of published work, heterogeneity between the texture features calculated under the same conditions but with different software. Standards for the comparison and

evaluation of the different software code used for the computation of the texture features are required.

- **Manual vs Automatic contouring of the ROI:** since tumors are highly-complex structures with irregular shapes and indistinguishable borders, the exact identification of the tumor volume is an extremely demanding task, if not impossible. Tumor segmentation can be performed either manually, semi-automatically, or automatically. Current data suggests that semi-automatic segmentation tend to provide more stable and reproducible results (higher robustness of the texture features), while automatic segmentation is the less reliable of the three options – pretty much an expected conclusion since computer-aided diagnosis (CAD) systems had failed in the past [37,39].

- **Image registration and contour propagation:** in order to accurately evaluate the progression – either that means growth or shrinkage of the tumor – of the disease in patients and the results of the therapeutic scheme (radiotherapy, chemotherapy), a reliable contour propagation is required. The right choice of a deformable image registration method is crucial since texture analysis performed in "non-treated" properly volumes can introduce several errors [37].

1.4.3.2 Image-Specific Factors

Besides, the general factors which concern the mathematical formulas and the parameters used – the way of calculation – for the computation of the texture features, there are other factors which may influence the computation process – unique for each modality (CT, MR, PET) used for the image acquisition. CT and MR techniques are usually incorporated to assist the quality resolution of PET images (e.g. attenuation of the sharp edges of the PET images, smoothening of the image) and increase PET's imaging accuracy, specificity, and sensitivity. Most of the time, texture analysis is performed only in the PET images.

1.4.3.2.1 CT-Specific Factors

- **Presence of metal artefacts:** it is common for CT images – especially in the head and neck region, usually because of the dental fillings – to observe texture features alterations due to metal streaks. Even though effective methods for the reduction of metal artefacts do exist, they could introduce artificial texture and modify the image's information. Thus, meaningful information may be lost or altered.

- **Noise & Blurring:** homogenous acquisition protocols usually reduce significantly the impact of errors due to noise and blurring. However, when they are not applied, the computation of texture features is heavily affected (resulting in features with significantly lower robustness).

- **Image reconstruction algorithms:** reconstruction algorithms in contrast-enhanced CT (CECT) images are reported not to or minorly affect the texture analysis process.

- **Contrast injection on CECT:** currently, there is not enough data to make safe and reliable conclusions regarding the dependency of contrast injection time, feature reproducibility between CECT scans, and whether CECT images could prove useful. Recent evidence is discouraging, as according to He et al. [40] conventional CT images were found to contain more meaningful information than CECT images [37].

1.4.3.2.2 MR-Specific Factors

- **Image acquisition protocols:** recent evidence suggests that texture features are extremely sensitive to acquisition parameter variations (especially with high spatial resolution) – as spatial resolution remains sufficiently high, it has little to no effect at pattern discrimination. However, more research is required.

- **Non-quantitative value of pixel intensities:** pixels in MR images don't have quantized values like in CT and PET imaging, application of pre-processing steps (e.g. correction of magnetic field inhomogeneities, intensity normalization) is required to overcome this limitation.

- **Robustness of parametric maps:** tissue heterogeneity can be measured in MR images using parametric maps generated by the use of model-fitting algorithms. This is a newly emerging field, therefore, only a limited amount of works is available and no conclusions can be made. Reproducibility, repeatability, and robustness of the texture features are yet to be assessed – current evidence suggests that reliability of the features is closely associated with the robustness of fitting of the model-fitting algorithm used [37].

1.4.3.2.3   PET-Specific Factors

- **Image acquisition & reconstruction protocols:** more than any other imaging modality, image acquisition, and reconstruction protocols heavily affect the texture features, with just a limited number of features proving to be robust. Time-of-flight, point-spread function, iteration number, and width of the Gaussian filter (reconstruction settings) have been reported to have a similar impact on the texture features, while grid size has a larger influence. Moreover, according to van Velden et al. [41], changes in delineation are way more influential than changes in reconstruction [37].

- **Lesion size:** scan time and lesion detection ability are linked in PET imaging, as longer scan times enable the detection of smaller lesions (for a typical 2 to 4 minutes' scan, lesions with a size smaller than 4.95 mm, it is likely that they would not be detected) [43].

- **Patient size:** patient size can have an impact on image acquisition, as bigger-size patients have an increased chance of generating scattered electrons which may end up alternating the PET image [37].

- **Image smoothing & Quantization:** PET images are full of noise, without the application of specialized models (image smoothing quantization models), the texture features are characterized by low robustness and poor reliability – making it impossible to perform texture analysis. The correct application of these models features the accurate estimation of the relationship between the image intensities and noises variances (in order to evaluate the corresponding parameters) [45].

- **SUV discretization:** SUV discretization criteria play a major role in the computation and interpretation of the texture features (as mentioned before). Therefore, more research is required to conclude in a standardized and widely-acceptable methodology [43].

- **Pre-processing step:** Hatt et al. [44] reported that the incorporation of tumor delineation and partial volume correction did impact in a positive manner the robustness of certain texture features [37].

1.4.4 Radiomics

Radiomics is a newly emerged field of research that takes advantage of certain quantitative metrics – the radiomic features – within the medical images. Radiomics are a continuum of computer-aided diagnosis and detection (CAD) systems with the main difference that CAD systems are stand-alone, while radiomics are used as decision-assisting tools. The field is based on the concept: "there is a lot of information hidden in the medical images which are cannot be perceived by the human eye and need to be extracted". The radiomic features are deriving from the texture analysis of the medical images, as mentioned above. Radiomics are attempting to link biological, biochemical or/and genomic characteristics of a pathological condition with certain quantitative metrics – they mostly find application in the field of oncology, whereas radiomic data (texture features capturing tissue and lesions properties and heterogeneity) may be combined with histologic, genomic, demographic, or proteomic data (under the supervision of machine learning or artificial intelligence (AI)) to lead in clinically useful

conclusions. Radiomics is a hybrid field (medicine and informatics), therefore, one of the main challenges is the lack of communication between the scientists of the two fields [46,47].
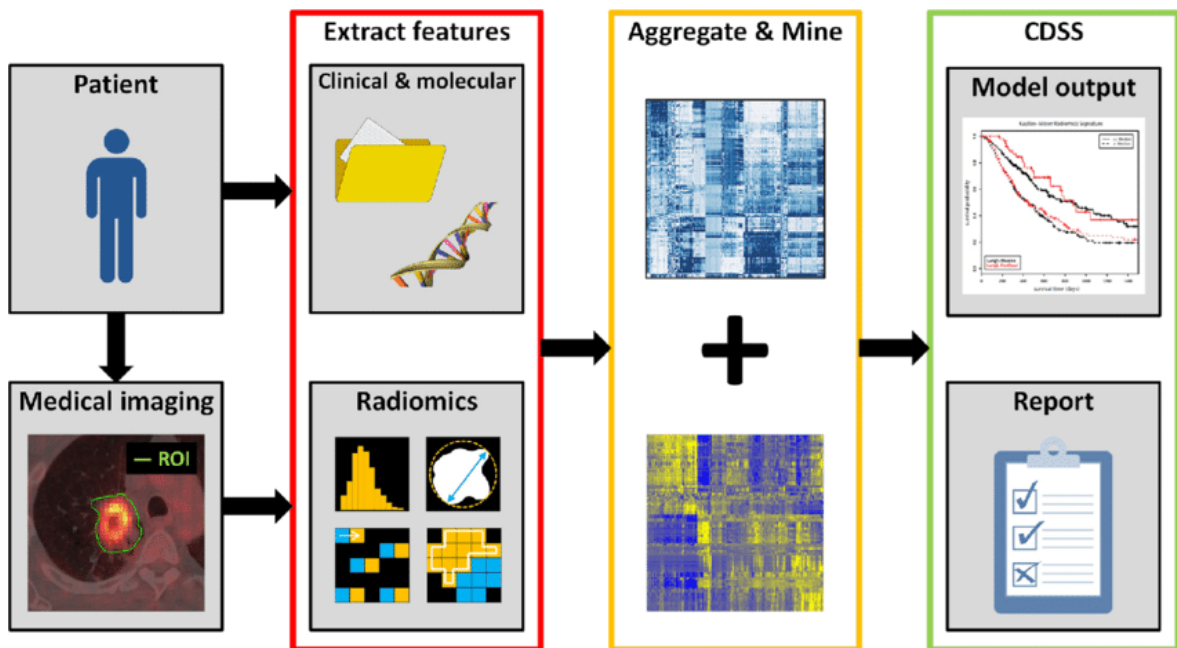
Figure 11. Display of Radiomics' methodological processes from the image acquisition to the data mining and assessment of clinical utility [70].

1.4.4.1 Radiomics in Oncology

 As mentioned before, Radiomics mainly find application in the field of oncology. Their main goal is to capture biological tissue and lesion characteristics (such as intra-tumoral heterogeneity and tumor's shape, but not limited to them), link them with certain texture features (quantitative features), and monitor their changes over time (e.g. during treatment or surveillance). Intra-tumoral heterogeneity is a well-known prognostic index – data derived from genomic studies – which has been linked with overall survival (OS), while it is being used extensively in cancer management and treatment planning. Even though tumor heterogeneity can already be measured through biopsy, the information provided is limited. Biopsies are providing meaning information only for a limited (usually

quite small) portion of the total tumor and just at a single anatomic site, while Radiomics are capturing the heterogeneity across the entire tumor – some studies have suggested that certain radiomic features may measure the heterogeneity even at a cellular level. Radiomic features have been associated with a variety of biological properties, including tumor aggressiveness, prediction of clinical response of a certain therapeutic scheme, genomic, transcriptomic, and proteomic data [46].

Radiomic data are derived from the texture analysis done in images, coming from different modalities (including PET, CT, and MR – PET images are of increased interest as they also provide metabolic information (metabolic imaging)). Radiomics are based on the concept that the data are mineable. Thus, if large datasets (the population-imaging approach which evaluates either unstructured data from different imaging modalities acquired for a very certain but unrelated diagnostic purpose or data derived from a single imaging modality but performed in a large cohort for a multicentric longitudinal observational study) are processed properly, newly, never-shown-before predictive markers and patterns of disease evolution, progression, and treatment response may come up. The upcoming radiomic data may be combined with histologic, laboratory, clinical, and genomic data using unsupervised machine learning – the real potential of Radiomics cannot be reached without the incorporation of the AI and machine learning which will evaluate hundreds of texture features to identify the very few that will be linked with certain biological properties [46].

1.4.4.2 Potential Clinical Applications of Radiomics

- **Radiogenomics:** is a subfield of Radiomics which attempts to link radiomic data with specific genomic patterns. Radiogenomics main goal is to define the genomic profile of cancer patients by using already existing data (imaging data) without submitting the patient to additional tests. One of the fields' limitations is the need for identification of both "valuable" – radiomic data which can be correlated with a gene mutation or expression status – and "invaluable" – radiomic data which have no genomic value – radiogenomic data, which may result in a better

prognosis. Radiogenomics have attracted the attention of researchers working with NSCLC patients. Nair et al. [49] conducted a clinical study and were able to link several radiomic features with four genes and OS. Furthermore, Yip et al. [50] in another clinical study, reported that 8 texture features were correlated with epidermal growth factor receptor mutation status, while one GLCM feature was linked with positive mutation status [46,47,48].

- **Clinical outcome prediction:** even though Radiomics are based on the analysis of huge datasets, they are paired with precision medicine (unique and custom treatment for every patient). Solid tumors usually prove to be different at both phenotypic, physiologic, and genomic levels, while they also tend to evolve – leading to intra-tumoral heterogeneity. This heterogeneity either intra-tumor or between tumor sites seems to be the main reason for treatment failure. Hence, early assessment of response to treatment and prediction of OS are crucial for treatment management and future monitoring, which will enable patient stratification. Arshad et al. [51] conducted a clinical study and were able to correctly predict the OS of NSCLC patients. Furthermore, Jiang et al. [52] in another clinical study on patients with gastric cancer, reported that the radiomic score had a high predictive value in terms of OS and disease-free survival [46,47].

- **Radiation-induced effects on normal tissues:** an infamous application of the Radiomics, is the predictive value of certain texture features in terms of identifying radiation-induced damage on organs (research has been focused on the lungs and the parotid glands) and tissues at risk. Cunliffe et al. [53] conducted a clinical study on patients with esophageal cancer and were able to link texture variations to the radiation dose, while 12 radiomic features were correlated with the development of radiation pneumonitis – a common side-effect of lung irradiation in this type of cancer patients [37].
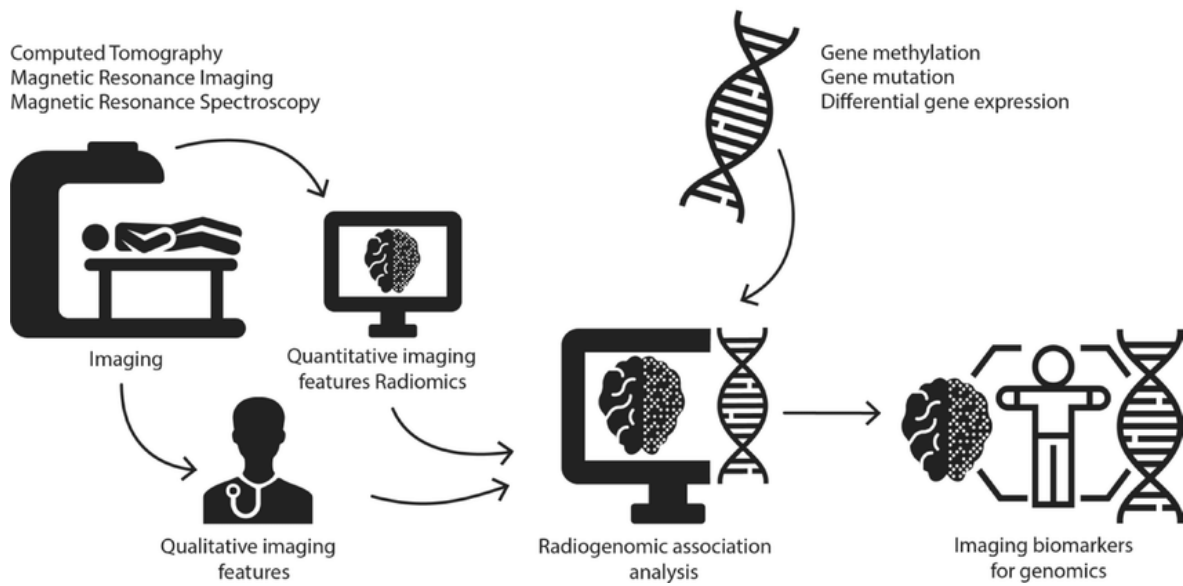
Figure 12. Overview of the research approach of Radiogenomics [71].

1.4.5 Challenges & Limitations of Texture Analysis

- **Standardization of the protocols:** there is an increased need for protocol standardization in terms of data acquisition, image acquisition, and reconstruction protocols. Additionally, the sample size of both the radiomic features computed in every image should be defined and the patients participating in the study – in most cases, there are more texture features than patients which may result in a high number of false-positives.

- **Standardization of the statistical analysis:** texture analysis usually provides an immense amount of complex data. To avoid the over-fitting of the radiomic features, a reduction in the number of features is required. Therefore, the need for the build of a classifier or a prediction is imperative.

- **Study design:** most of the time, texture analysis is used as a proof-of-concept in retrospective and preliminary studies. Thus, it lacks complete control over data acquisition and management – negatively affecting the reproducibility and robustness of the results. Another issue is the need for generalization and validation of the radiomic signature found in the assessed patient population –

validation of the radiomic signature on an external group of patients must be established.

- **Biological interpretation of the data:** the biological origin of the vast majority of the texture features is still unclear or poorly understood. More studies should focus on the bridging of the radiomic data and tumor's biological properties or proteomic/genomic data – intra-tumoral heterogeneity has already been linked with pathophysiological properties (identified by genomic analysis), suggesting that heterogeneity features may show a degree of correlation with cell cycling pathways [37].

Figure 13. Multimodal imaging modalities for small-animal imaging [72].

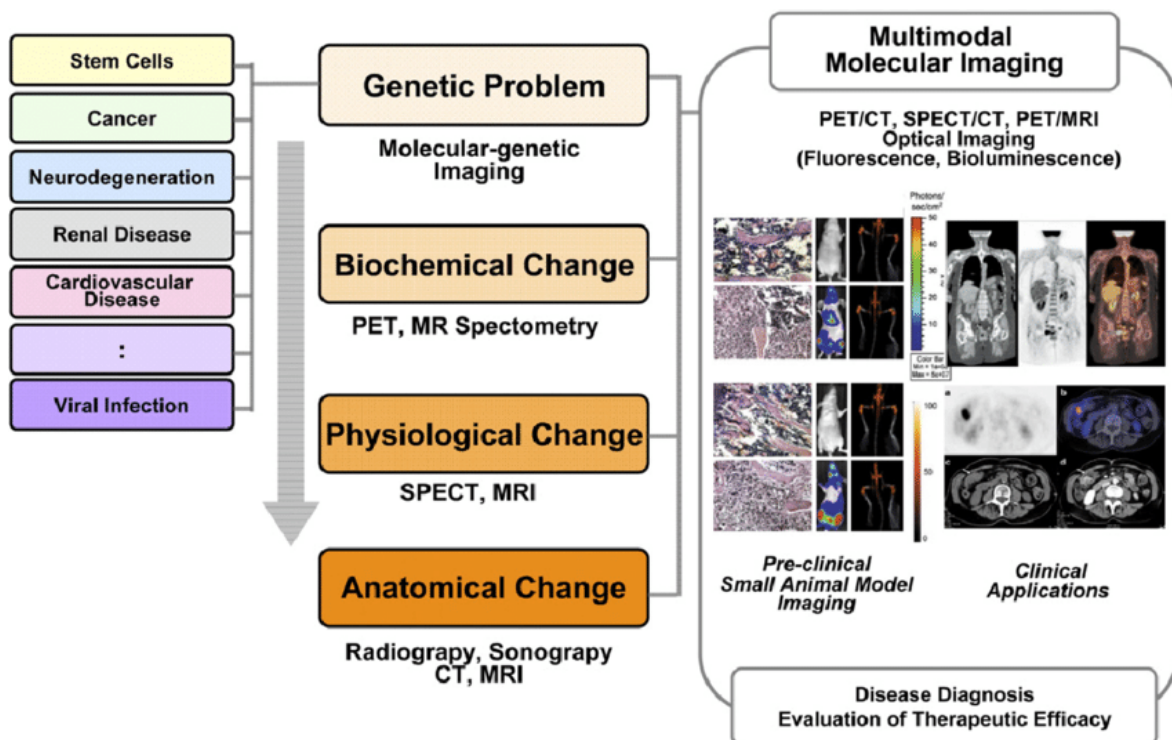### 1.4.6   Importance of Small-Animal Imaging

There is an increased need for small-animal imaging, as the research's foundations are based on animal models. The animals of choice are usually mice and rats due to their

ability to host a huge variety of human diseases – ranging from different cancer types (e.g. melanoma) to diseases of the immune system (e.g. immunosuppression). Small-animal PET imaging emerged in the mid-1990s, since then it has been linked with the modern biomedical as it can provide information about: the dimensional distribution of a radiopharmaceutical (as a function of time in a live subject), the dynamic biodistribution of a radiopharmaceutical (or of a labeled compound), and the pharmacokinetic profile of a newly derive drug – while it allows a single animal to be tested a lot of times through the course of the study, it reduces the cost of laboratory animals and animal testing and provides quantitative information. Small-animal studies may also provide information regarding the dose, route of administration, metabolism, effectiveness, and toxicity of a newly developed drug. One of the main advantages of the small-animal PET is that (similar to the PET imaging for humans) it can be computed with other imaging techniques (such as CT or MR) to improve the quality of the generated image and provide information about the physiological properties and molecular abnormalities caused by disease (in quantitative form) [57].

As mentioned before, the PET, CT, or MR data acquired is quantitative and can be measured. Thus, small-animal imaging can be used to assist in the development of the field of Radiomics and Texture Analysis. Indeed, there are a lot of animal studies [58,59,60,61] (the current one included) employing animals to identify radiomic features that could serve as potential biomarkers in human diseases. In a recent study, Veres et al. [61] injected a ^{99m}Tc-albumin nanoparticle into mice with liver cancer and reported that they were able to identify cold spot cancer lesions, which were at first invisible in the SPECT scan, from the skewness of the activity distribution in the image.

2. Materials & Methods

2.1 Small-Animal PET/CT Scanner Features

 The Mediso NanoScan®PC (PET8/2) – a small-animal PET/CT scanner with two rings (containing eight detector modules each) – was used for conducting the imaging experiments. All the detector modules are identical; consisting of 29 x 29 crystals – LYSO crystal material, size: $1.51 \times 1.51 \times 10$ mm^3. Position sensitive photomultiplier tubes (256 channels each and associated front-end electronics) are attached to the modules. Scanner's field of view (FOV) consists of the axial, which is 98.6 mm, and the transaxial, which is 80 mm. The acquisition of the coincidence counts is performed in the list mode format. The data is sorted into 3D lines of response (LOR). Image reconstruction is performed using a Tera-Tomo 3D algorithm – a total image reconstruction algorithm, which incorporates the OSEM algorithm, the TV term (level of regularization), and the modeling of detector response, positron range, gamma attenuation, and scatter effects [73, 74]. Spatial resolution at 1 mm off the center of the scanner in the radial, tangential and axial directions is 0.8 mm, 0.8 mm, and 0.85 mm, respectively. Reconstructed image's voxels have a size of $0.4 \times 0.4 \times 0.4$ mm^3. Corrections related to dead time, decay, scatter, attenuation, axial sensitivity, and normalization are applied to the "raw" PET data (not smooth, not continuous, sharp edges). Attenuation correction of the emission data is done using a CT scanner. Emission data were recorded for 20 minutes using the default energy window of 400 – 600 keV. Tera-Tomo 3D reconstruction algorithm offers a selection of the subsets/iteration combination, as well the selection of the three levels of regularization: low ($\beta=10-4$), normal ($\beta=5\times10-4$), and high ($\beta=10-3$). The total level of regularization is used for the handling of the ill-conditioned nature of the EM reconstruction algorithm [73].

2.2 Positron Emitting Tracer

A well-studied and widely used positron-emitting tracer; 18F-FDG (Viokosmos SA, Lavrio, Athens) was used for the conduction of the experiments. 18F-FDG allows for the simultaneous acquisition of PET and CT data. It is an analog of glucose and can provide valuable information based on glucose uptake and glycolysis – identifying lesions with increased metabolism. 18F-FDG has unique physicochemical properties; a half time of 109.8 minutes, a β+ yield of 0.97, a mean β+ energy of 0.25 MeV, and a mean range of 0.64 mm [75].

Figure 14. Images from the front and anterior side of the Mediso NanoScan®PC. In the first image, the IQ phantom which was used in the studies, can be seen.

2.3 Image Quality (IQ) Phantom

IQ phantom is a 63-mm-long poly(methyl methacrylate) (PMMA) fillable cylinder with a diameter of 33.5 mm and an inside diameter of 30 mm. Three distinct regions can be spotted – each one simulating different imaging conditions (cylinders part is characterized by a high ΔSUV (SUVmax-SUVmin – >1.5 MBq) simulating heterogeneous tumors (in terms of uptake), homogenous region is characterized by lower ΔSUV (~0.10 MBq) simulating highly homogenous tumors, while the fillable chamber part with the two cylinders simulates a tumor with necrotic areas – treated tumor). The first part of the phantom is

20 mm long, and it has 5 fillable rods of different diameters – 1 mm, 2 mm, 3 mm, 4 mm, 5 mm. The second part of the phantom is a uniform and homogeneous 33-mm-long cylinder with an inner diameter of 30.5 mm. The third part of the phantom is a fillable chamber with two 14-mm-long cylinders, with inner diameters of 30 mm. The one of the two cylinders contains non-radioactive water, while the other one air [76].

2.4 Animals

A total of twelve (n=12) mice were maintained in the animal facility of the Biomedical Research Foundation of Academy of Athens (BRFAA). The mice used in the experiments were all male, 8 to10 weeks old. They were split into two groups: Treated (T) and Non-Treated (NT). The mice of the Treated group received treatment on the 10th day, while the Non-Treated mice did not. All the procedures were in accordance with the institutional guidelines and were approved by the Greek Federal Veterinary Office [77].
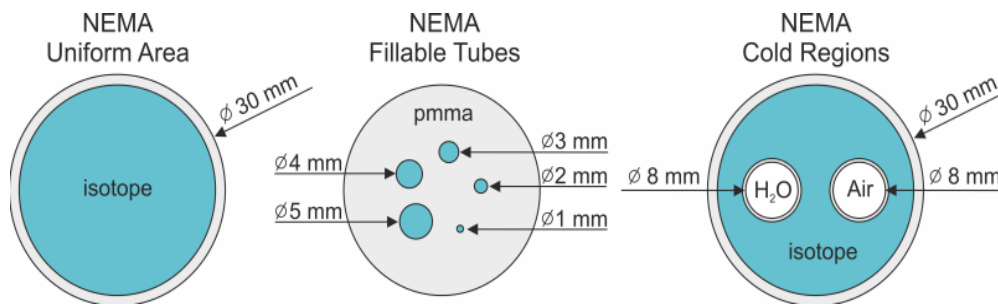

Figure 15. Display of the NEMA IQ phantom.

2.5 PET Acquisitions of the IQ Phantom

 The activity of 18F-FDG used in IQ phantom's acquisitions was 3.7 MBq – following the NEMA standards [78] – and the duration of the PET acquisition was 20 minutes.

2.6 PET/CT Imaging of the Animals

The mice (both treated and non-treated) underwent a fasting period of 12 hours before the imaging procedure. The PET acquisitions were performed under anesthesia, using 2% isoflurane in 0.8 L/min O_2 – with continuous monitoring of the vital signs during the anesthesia period. The temperature of the mice was sustained at 36°C during the imaging process; thus, mice were placed onto warm beds. The 18F-FDG was injected in the lateral tail vein. Similar to the PET acquisitions of the IQ phantom, the procedure lasted for 20 minutes, but an uptake time of 60 minutes was preceded [77].

2.7 Image Reconstruction Parameters

The image reconstruction parameters used in the IQ phantom experiments were: 4 subsets; 14 iterations (for 4 subsets); high regularization (HR) levels. Even though the scanner is able for reconstruction with 1, 2, 3, and 4 subsets, because of the eight detector modules of the PET scanner, reconstruction was performed only for 4 subsets. These parameters were chosen based on previous research experience [76], as they were reported as optimal for the Mediso NanoScan®PC (PET8/2).

The image reconstruction parameters, for the reconstruction of both PET and CT data, used in the animal experiments were similar to the reconstruction parameters used in the IQ phantom study – 4 subsets; 14 iterations (for 4 subsets); high regularization levels [76].

2.8 Image Analysis

The image analysis was performed using LIFEx version 6.20 [79]. Five regions of interest (ROIs) were set in each set of images of the IQ phantoms. The four ROIs were set onto the middle of the fillable rods (with a diameter of 2 to 5 mm) of the first part of the phantom, while another one was set onto the uniform area – second part – of the phantom, as it is

seen in **Figure 16.** Because of the extremely small size of the 1-mm-diameter rod, a ROI could not be set there – some textural indices couldn't be calculated. Regarding the image analysis of the mice, one ROI (the tumor area, as it is seen in **Figure 17.**) was set manually in each set of images, using a 3D selection tool of the image analysis software.

First-Order features, Second-Order features, and Higher-Order features were calculated for all the ROIs in all cases (for a total of 20 different IQ phantom acquisitions). First-Order features include Histogram indices, Shape indices, Conventional indices, and Discretized (DIS) indices. Second-Order features include Grey Level Co-occurrence Matrix (GLCM) indices. Higher-Order features include Neighborhood Grey Level Difference Matrix (GLNM) indices, Grey Level Zone Length Matrix (GLZLM) indices, and Grey Level Run Length Matrix (GLRLM) indices. The texture indices calculated can be seen in **Table 1.** – First-Order features provide information about the distribution of the grey levels in the image, while the Second-Order features provide information about the spatial distribution (e.g. relative position) of the grey levels in the image [79].

The texture indices were calculated for a total of three different bin sizes – three different numbers of grey levels (noGL) [43,80]. The bin sizes were: 0.317, 0.157, 0.078 which correspond to 64, 128, and 256 number of grey levels, respectively – 32, 64, 128, and 256 noGL and their corresponding bin sizes are already well-studied, however results from different studies vary (some studies [43] suggest bin sizes in the 0.05-0.1 range generate more robust features, while others [80] propose a bin size of 0.10 and 0.25 as the optimal). At first, the features were also computed for the 32 number of grey levels, however a lot of outlier values were noticed – thus, the results were considered unreliable. This was done to find out whether the SUV discretization has an impact on the computation of the texture features, as well as to find out which number of grey levels is the optimum for medical analysis, if any.
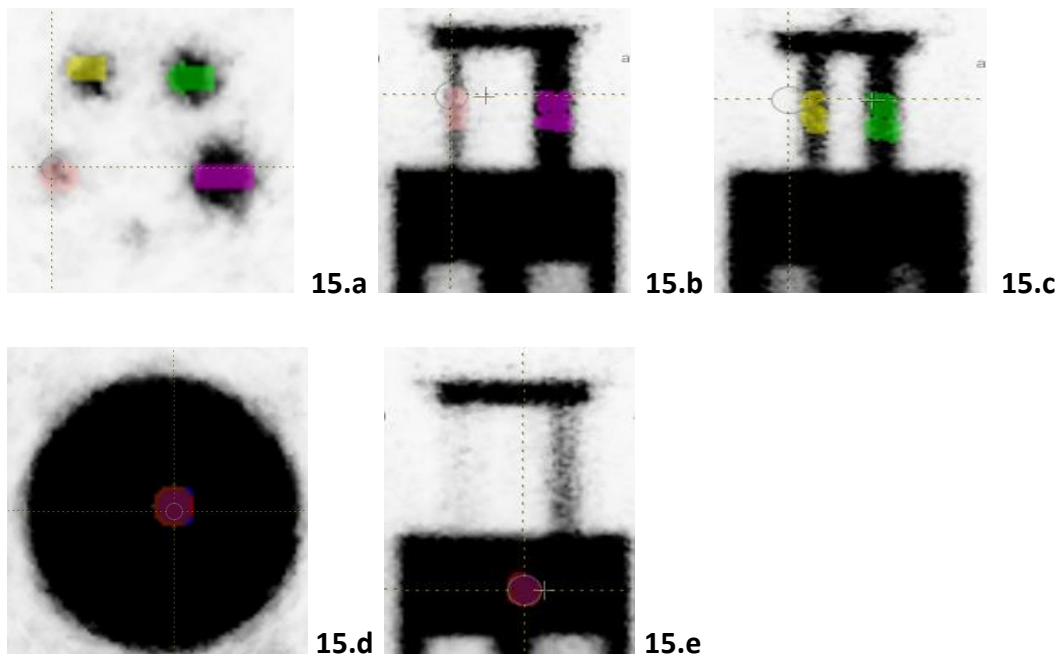
Figure 16. The ROIs used in the texture analysis of the IQ phantoms. Four ROIs were set onto the fillable rods: a ROI with a volume of 0.045 cm^3 and 696 voxels was set on the 2-mm-diameter rod, a ROI with a volume of 0.059 cm^3 and 916 voxels was set on the 3-mm-diameter rod, a ROI with a volume of 0.091 cm^3 and 1418 voxels was set on the 4-mm-diameter rod, and a ROI with a volume of 0.106 cm^3 and 1663 voxels was set on the 5-mm diameter rod. The ROIs of the cylinder can be seen in the images **15.a** (axial view), **15.b** (coronal view), **15.c** (sagittal view). One ROI was set onto the uniform area of the phantom, with a volume of 0.06573 cm^3 and 1027 voxels. The ROI has a circle-like shape (when having voxels e.g. squares, it is impossible to achieve perfect circles) and can be seen in the images **15.d** (axial view), **15.e** (coronal view). The size of the ROIs was determined based on previous research experience [81] – even though the voxel size and the size of the ROIs were different in the referred paper, similar criteria (each ROI sound contain enough voxel for the regular computation of the texture features, different sizes of ROIs were used to assess whether the size affects the computation process, sizes of the ROIs of the cylinders are similar to the size of the ROI of the uniform area to assess the impact of the different imaging conditions) were used.

Certain features (Entropy, Energy, and Contrast) are computed for more than one class (order) of features. First-Order Entropy provide information about the randomness of the intensities of grey level into the histogram, while Second-Order Entropy provide information regarding the relationship (whether are random or not) of two separate

events which are the relative position and the distribution of the grey level intensities in the 3D matrix [82]. Same applies for First-Order and Second-Order Energy but this time in terms of uniformity. In case of Contrast, Second-Order Contrast provide information about the local variations of grey level intensities, while Higher-Order Contrast provide information about the variations of grey level intensities between different neighborhoods (composed of at least 27 voxels) of voxels [79].

16.a

16.b
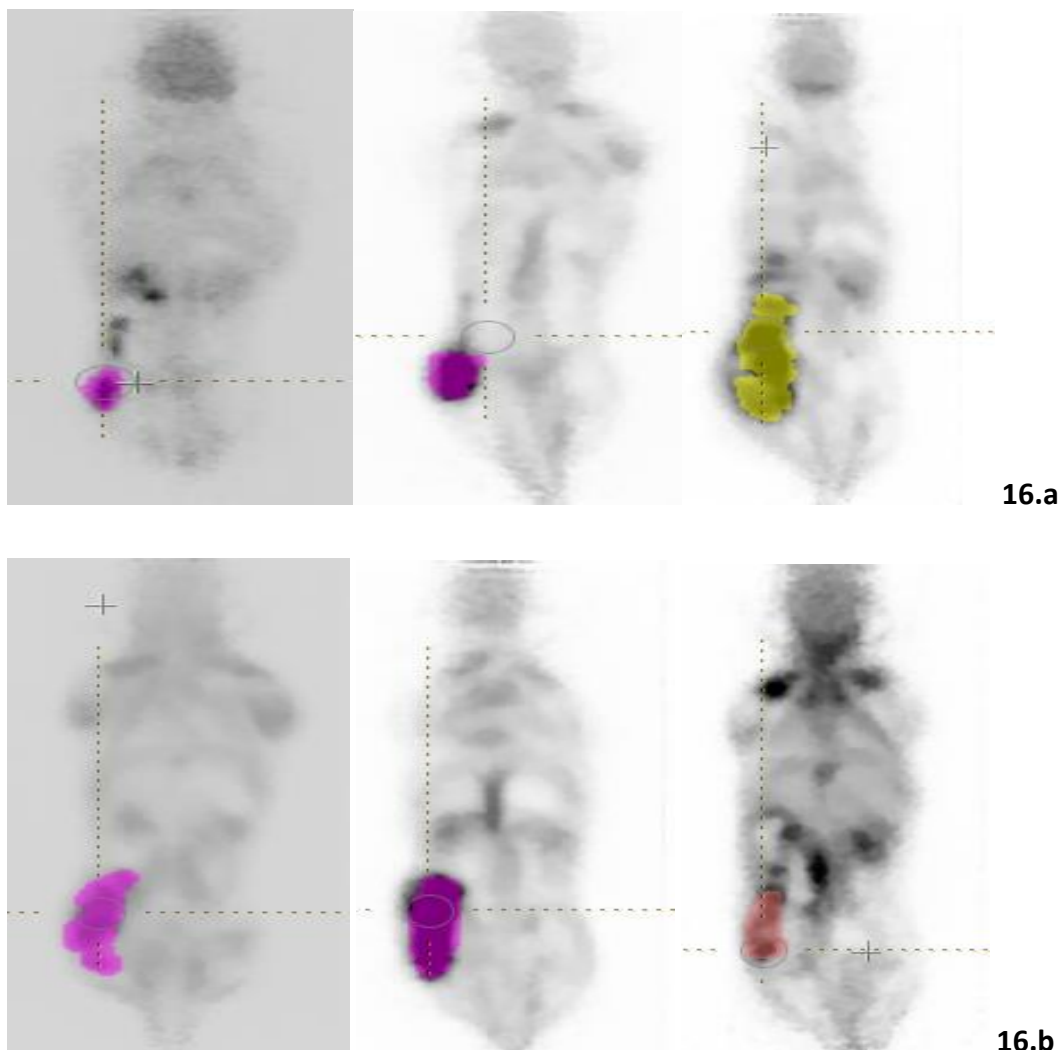
Figure 17. The evolution of a tumor in a non-treated **(16.a)** and a treated **(16.b)** mouse, over the course of 7, 10, and 14 days (from left (7 days) to right (14 days)). **16.a** The evolution of the tumor in a non-treated mouse over the period of 7, 10, and 14 days. It can easily be seen that the tumor grows as the time passes by. **16.b** The evolution of the tumor in a treated mouse over the period of 7, 10, and 14 days. Tumor shirnks in size from day 10 to day 14.

Table 1. Total texture indices (51 texture indices) calculated in the ROIs. Table **1.a** shows the total of the First-Order features, **1.b** the total of Second-Order features, and **1.c** the total of the Higher-Order features.

First-Order Features

Histogram	Shape	Convetional	Discretized
Skewness	Volume (mL)	SUVmin	SUVmin
Kurtosis	Volume (vx)	SUVmean	SUVmean
ExcessKurtosis	Sphericity	SUVmax	SUVmax
Entropy_log10	Surface (mm^2)	TLG (mL)	TLG (mL)
Entropy_log2	Compacity		
Energy			

1.a

Second-Order Features

Grey Level Co-Occurrence Matrix (GLCM)
Homogeneity
Energy
Contrast
Correlation
Entropy_log10
Entropy_log2
Dissimilarity

1.b

Higher-Order Features

Grey Level Run Length Matrix (GLRLM)	Neighboorhood Grey Level Difference Matrix (NGLDM)	Grey Level Zone Length Matrix (GLZLM)
Short Run Emphasis (SRE)	Coarseness	Short Zone Emphasis (SZE)
Long Run Emphasis (LRE)	Busyness	Long Zone Emphasis (LZE)
Low Gray-level Run Length Emphasis (LGRE)	Contrast	Low Gray-level Zone Length Emphasis (LGZE)
High Gray-level Run Length Emphasis (HGRE)		High Gray-level Zone Length Emphasis (HGZE)
Short Run Low Gray-level Run Length Emphasis (SRLGRLE)		Short Run Low Gray-level Zone Length Emphasis (SRLGZLE)
Short Run High Gray-level Run Length Emphasis (SRHGRLE)		Short Run High Gray-level Zone Length Emphasis (SRHGZLE)
Long Run Low Gray-level Run Length Emphasis (LRLGRLE)		Long Run Low Gray-level Zone Length Emphasis (LRLGZLE)
Long Run High Gray-level Run Length Emphasis (LRHGRLE)		Long Run High Gray-level Zone Length Emphasis (LRHGZLE)
Gray Level Non Uniformity (GLNU)		Gray Level Non Uniformity (GLNU)
Run Length Non Uniformity (RLNU)		Zone Length Non Uniformity (ZLNU)
Run Percentage (RP)		Zone Percentage (ZP)

1.c

2.9 LIFEx Software

LIFEx [79] is a user-friendly medical analysis software (Osirix-compatible software [83]), which reads medical images (locally or over network) using a DICOM browser (only available for DICOM images) – LIFEx interface can be seen in **Figure 18.** It runs well both in MacOs, Windows and Linux. LIFEx offers a lot of special features, including a dynamic 3D-reconstruction based slice viewer, while it allows to the user to modify the computation parameters (e.g. bin size, number of grey levels, resampling method) of the texture features. A lot of research labs, radiology and medical departments are using it for texture feature analysis and tumor identification purposes. Medical images processed by LIFEx, most of the time derive from PET or CT imaging and rarely from MR or US imaging. A huge variety of texture indices (over 86 different texture features) is given, which are associated with the number of grey levels (existing in the image) and their distribution (spatial (Second-Order features) or not (First-Order features)), frequency, or other variable (Higher-Order features). The computed features are split into three main categories (based on their way of calculation): First-Order features (Conventional, Discretized, Shape, and Histogram features), Second-Order features (GLCM), and Higher-Order features (GLRLM, NGLDM, and GLZLM). These indices are computed from ROIs which can either be imported from an already existing file or set manually. Results are exported to excel or cvs format files. First-Order features are computed from the distribution of the grey level intensities into the histogram, while Second-Order and Higher-Order features are calculated for a total of 13 directions in the 3D matrix (as seen in **Figure 19.**) – the appearing index value of each feature is the average of the values of the 13 directions [79].
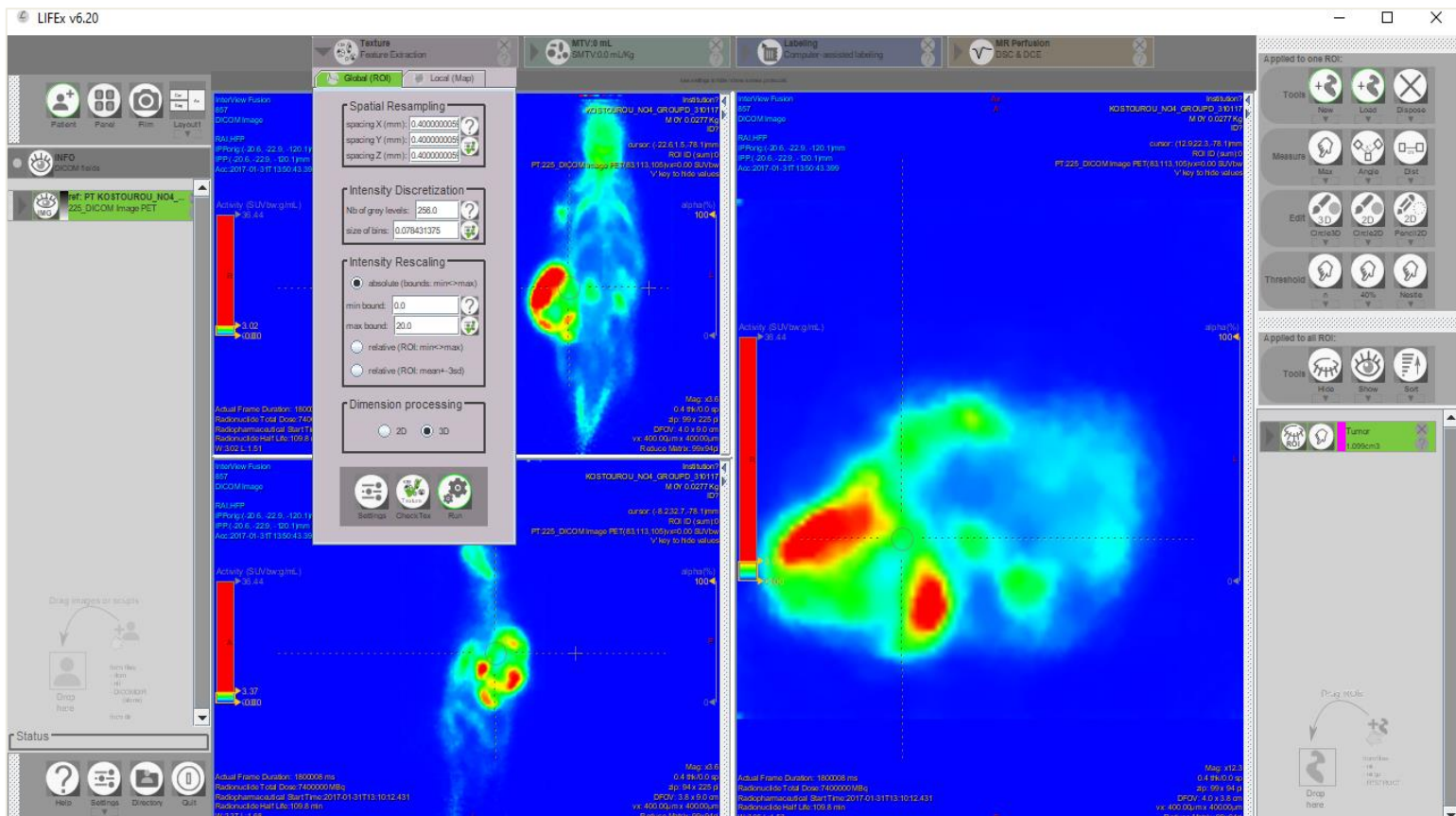
Figure 18. Display of the LIFEx version 6.20 interface.

Furthermore, LIFEx has contributed positively in the field of the nuclear medicine with the creation of databases and public data banks (PET and CT databases, an MR database is currently "under construction") – such data enable an accurate characterization of the variability of different texture indices in a given imaging modality as a function of the scanner and the imaging protocol (all the users have the ability to participate in that project) – to better help with the interpretation of the results [79].
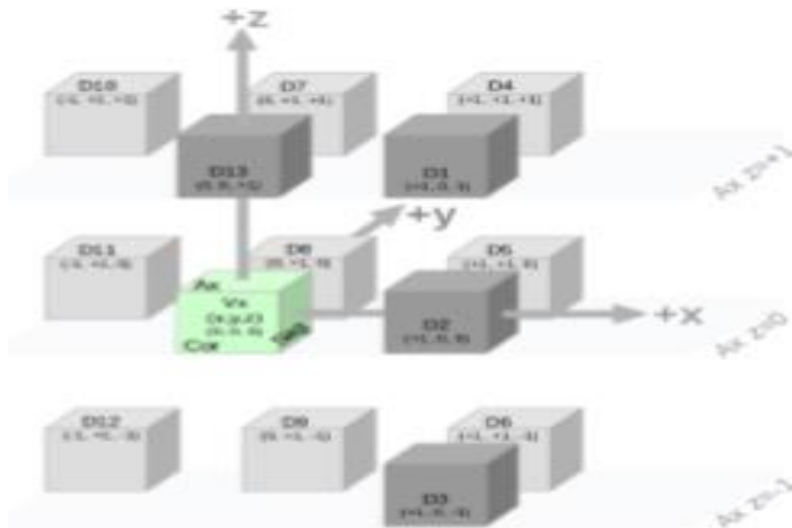
Figure 19. Display of the 13 directions (gray boxes) for which the radiomic features are computed for, in the 3D matrix – 3D Voxel (green box) [79].

2.10 Statistical Analysis

Statistical analysis was performed five different statistical tests: Interclass Correlation Coefficient (ICC), Spearman's Rank Correlation Coefficient (SRCC), One-Way Analysis of Variance (one-way ANOVA), Least Significance Difference (LSD). Coefficient of Variation (CV) and percent difference (PD) between the values were also calculated. COV and PD were calculated to check the liability of the indices.

2.10.1 Interclass Correlation Coefficient

ICC test was incorporated to determine whether the number of grey levels chosen is affecting the numeric value of the texture indices and which is the correlation between the number of the grey levels, if any. ICC values range from -1 to +1, where -1 indicates a perfect negative correlation between the results (as the number of grey levels is increasing the numeric value of the indices is decreasing), +1 indicates a perfect positive correlation (as the number of grey levels is increasing the numeric value of the indices is

increasing too), while 0 indicates no correlation (the number of grey levels is not affecting the numeric value of the indices). To determine the level of correlation indicated by the ICC values, the guidelines published by Koo and Li [84] were used: values lower than 0.5 (but higher than 0) were considered to show poor correlation (similarly values greater than -0.5 signified poor negative correlation), values ranging from 0.5 to 0.749 were considered to show moderate correlation (similarly values ranging from -0.5-(-0.749) signified moderate negative correlation), values ranging from 0.75 to 0.899 were considered to show good correlation (similarly values ranging from -0.75-(-0.899) signified good negative correlation), while values greater than 0.9 were considered to show excellent correlation (similarly values lower than -0.9 signified excellent negative correlation). The mathematical formula used for the computation of the ICC was:

$$ICC(X,Y) = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

where ICC is Interclass Correlation Coefficient, and x and y are the numeric values of the compared array, while $\bar{x}$ and $\bar{y}$ their mean values.

2.10.2 Spearman's Rank Correlation Coefficient

SPCC was incorporated to determine whether there is a rank correlation between the numeric values of the texture indices in the different number of grey levels – in every single feature value of each ROI for each acquisition (20 values total for each feature on each ROI), a rank is applied, then the ability of each feature value to retain its rank over the different number of grey levels is examined. SPCC takes its highest values (+1) when the rank of the observation between the variables is not changing and its lowest values (-1) when the rank of the observation between the variables is fully opposed. Similar to ICC,

the same guidelines were used for the evaluation of the results. The mathematical formula used for SPCC was:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

where r_s is SPCC, d_i is the difference between the two ranks of each observation, and n is the number of the observations.

2.10.3 One-Way Analysis of Variance & Least Significance Difference

One-way ANOVA was incorporated to determine the statistical significance of the results. The alpha-value was set to 0.05 – alpha-value is the probability of rejecting the null hypothesis when considered true, it is set manually, and usually is 0.05. ANOVA was calculated for the total of the measurements at each number of grey levels for each radiomic feature in every region of interest. In the case of the animal studies, it was used to determine whether the treatment of the tumor has an impact on the numeric values of the radiomic features. A p-value (p-value is a calculated probability, depending on the mathematical formula which was computed from) lower than 0.05 (alpha-value) was considered enough to determine statistical significance.

In the cases that null hypothesis was rejected, an LSD test was incorporated. The purpose of the LSD, is to determine between which groups of noGL the statistical difference was noticed – a p-value lower than 0.05 does not necessarily mean that there is a significant difference between all three pairwise comparisons (64-128, 64-256, and 128-256 noGL). An LSD critical value for each radiomic feature was calculated, using the following mathematical equation:

$$LSD = t \sqrt{MSW \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

where LSD is least significance test, MSW is the mean square within the groups (calculated by the ANOVA table), N_1 and N_2 are the numbers of the observations of the groups 1 and 2, respectively, and t is the critical value from the t-distribution table. The LSD critical value was then compared to the absolute value of the difference of the mean values of the groups being-compared ($|\bar{x} - \bar{y}|$, where $\bar{x}$ and $\bar{y}$ are the mean values of the groups). If $|\bar{x} - \bar{y}| > LSD$, then the result was considered significant.

2.10.4 Coefficient of Variation

COV was calculated for each textural feature over the different number of grey levels (that were applied), using the following equation:

$$COV = \frac{SD}{\bar{x}} \times 100$$

where CV is coefficient of variation, SD is the standard deviation of the texture feature, and $\bar{x}$ is the mean value of the texture feature over applying different number of grey levels. COV was incorporated to determine the liability of the texture features (which is the variance of the numeric values of the texture features over different number of grey levels). COV values near to 0% indicate very small variance, meaning that the examined texture feature is stable at a given number of grey levels (meaning that it can be used in texture analysis, and may be a potential biomarker). To figure out the level of variance: COV values smaller or even to 5% signified a very small variance, values ranging from 5% to 10% signified a small variance, values ranging from 10% to 20% signified a moderate

variance, values ranging from 20% to 30% signified a large variance, while values greater than 30% signified that the given feature is heterogeneous.

2.10.5 Percent Difference

PD was calculated for each texture feature over the different number of grey levels (that were applied), using the following equation:

$$PD = \frac{(x - \bar{x})}{\bar{x}} \times 100$$

where PD is Percent Difference, x is the texture feature value of an IQ phantom at a given number of grey levels, and $\bar{x}$ is the average value of the texture feature at the given number of grey levels. PD was incorporated to determine the deviation of each texture feature value of each ROI at a given number of grey levels from its average value – in simple terms to check texture feature's liability over the different number of grey levels applied (low values of PD indicate small deviation/variance, thus the examined texture feature is stable at the given great level). To figure out the level of variance, the same principles as Galavis et al. [85] were used: PD values lower than 5% signified small variance, values ranging from 10% to 25% signified moderate variance, while values greater than 30% signified large variance.

3. Results & Discussion

 Applications of the texture analysis in medical images have been gaining a lot of attention, lately. In the present study, we investigated the reliability of the radiomic features in terms of ICC and SPCC. The liability of the features was also tested through the calculation of COV and PD, while the statistical significance of the results was concluded through one-way ANOVA and LSD tests. Data presented in the current study should be used as an evaluation tool for the assessment of the performance of the radiomic features over different numbers of grey levels (different bin sizes), as well as for the selection of certain 'well-performing' features – features with great repeatability, reproducibility, and liability, while being both robust to the bin size applied and able to provide valuable information to separate a diseased tissue from a recovering or a healthy one – to determine whether they could have a place in the clinical routine. This is not the first study that tests the clinical applicability of the radiomic features [41,43,85,89]. However, in the depts of our knowledge, this is the very first study that compares the results of the texture analysis done in a series of IQ phantom acquisitions to results deriving from preclinical data (Treated and Non-Treated mice with melanoma).

3.1 IQ Phantom Study

 ICC test is a non-parametric statistical test used to determine the correlation and the agreement of the radiomic features over different numbers of grey levels (noGL). All the features examined were found to have at least good reliability (>0.75) in all three pairwise comparisons, with most of them achieving excellent reliability (>0.90). Only a limited amount of features performed poorly, showing moderate (<0.75) to poor (0.50) correlation over the different noGL applied – these features were First-Order features (Discretized SUVmin, SUVmean, SUVmax, and TLG) and Higher-Order features (GLRLM SRLGLE and GLZLM SZLGLE, LZHGLE, and GLNU), indicating that some First-Order and Higher-Order features are heavily

affected by the noGL applied. The results of the ICC tests can be seen in **Figure 20.** More specific, 34 out of 42 features (80% of the features in total) show at least good reliability (>0.75), while 28 out of 42 (67% of the features in total) features show excellent reliability. ICC test was not calculated for the Conventional and Shape indices, as the total number of grey levels does not affect the computation of these features – they refer to innate properties and characteristics of the ROI or image (e.g. SUV activity, shape). Most of the features showing excellent reliability or better were deriving from the Second-Order features group (7 out of 7 GLCM features – 100% of the features), while First-Order and Higher-Order features showed similar performance in terms of total features exceeding the 0.95 mark of ICC test (6 out of 10 features for the First-Order features (60% of the features) and 14 out of 25 for the Higher-Order features (56% of the features)). These results (34 out of 42 features scoring over 0.8 – 80% of the features) come in agreement with the Leijenaar et al. [43] study in which 75% of the radiomic features investigated showed high reliability, leading us to the conclusion that the performance of the most radiomic features is not affected by the total noGL (bin size) – a conclusion that has already been reported in another Leijenaar et al. [89] study for clinical PET images.

 SPCC is another non-parametric statistical test used to determine the impact of the noGL on the ranking of the acquisitions (pairwise comparison of the phantom acqusitions between the different noGLs for each feature) – SPCC checks whether the rank of the acquisition is affected (changes value or not – e.g. rank 4 at 64 NOGL, rank changes to 9 at 128 noGL) by the noGL (basically inter-subject agreement of the results), an ICC test but instead of values ranks are incorporated. Rank repeatability of the features over the different bin sizes (64, 128, and 256 noGL) does not necessarily equal robustness (reproducibility, repeatability, and liability). All of the texture features were found to provide a reliable acquisition ranking over the different noGL tested – showing at least good rank reliability (>0.75), while most of them achieved excellent rank reliability (>0.90). The features which did not achieve
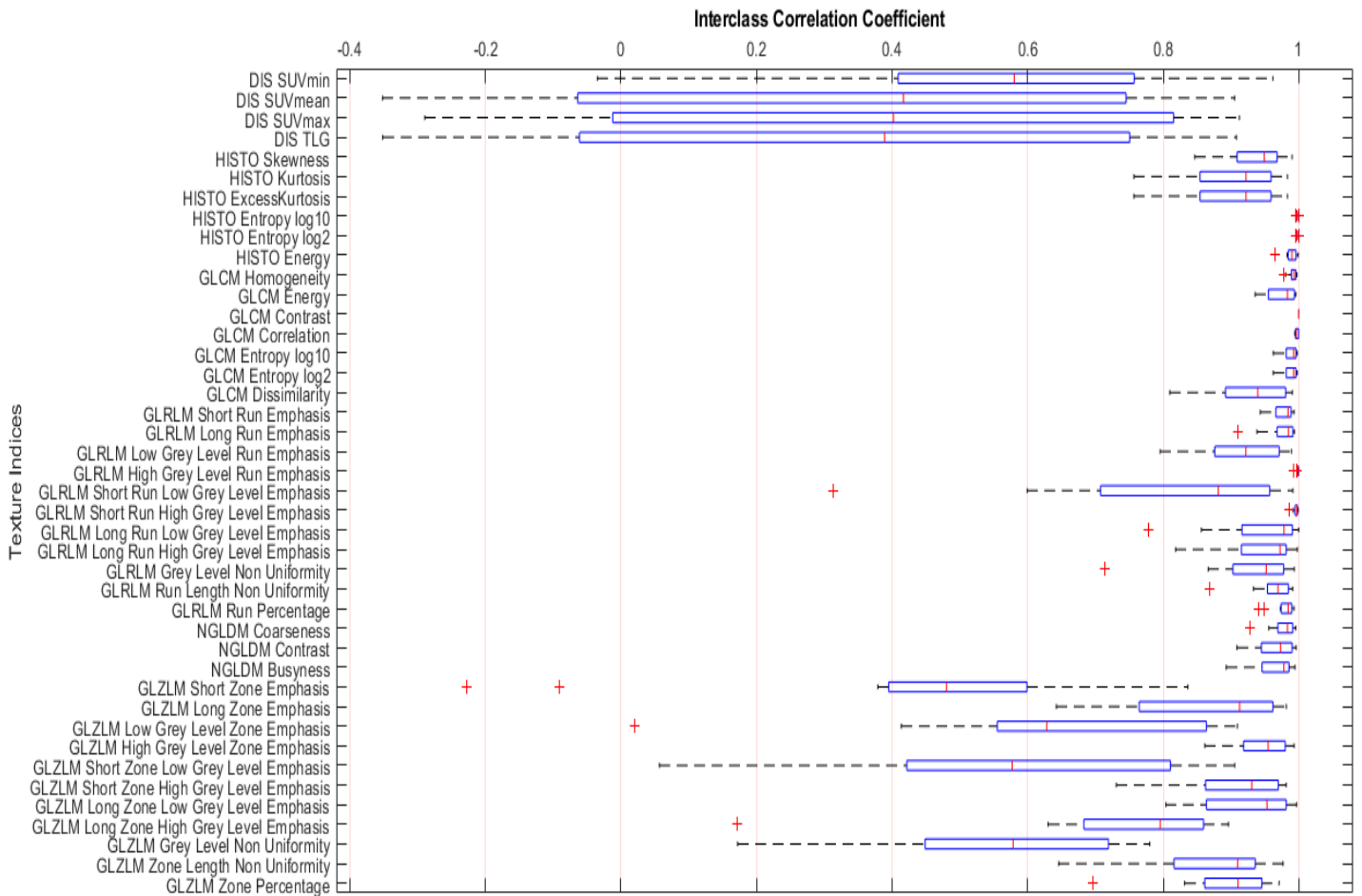
Figure 20. ICC medians of the 2mm-diameter, 3mm-diameter, 4mm-diameter, and 5mm-diameter cylinders. ICC values lower than 0.5 (but higher than 0)are indicating poor correlation (similarly values greater than -0.5 signify poor negative correlation), values ranging from 0.5 to 0.749 are indicating moderate correlation (similarly values ranging from (-0.5)-(-0.749) signify moderate negative correlation), values ranging from 0.75 to 0.899 are indicating good correlation (similarly values ranging from -0.75-(-0.899) signify good negative correlation), while values greater than 0.9 are indicating excellent correlation (similarly values lower than -0.9 signify excellent negative correlation). **Red cross:** outlier value, **Red vertical line:** median, **Blue Parallelogram:** range of values.

excellent but moderate (<0.75) rank reliability were Higher-Order features (GLRLM SRLGLE and GLZLM SZE, LGLZE, HGLZE, SZLGLE, SZHGLE, LZHGLE, GLNU, and ZLNU) – indicating that most of the GLZLM features are not able to provide reliable rankings when different noGLs are applied (thus, even the slightest of a

change in the acquisition and reconstruction processes of these features, may alter their prognostic (or any) value). The results of the SPCC tests can be seen in **Figure 21.** More specific, 36 out of 42 radiomic features (85% of the features in total) show good rank repeatability or better, while 24 out of 42 radiomic features (57% of the features in total) show excellent repeatability. Most of the features which exceeded the 0.95 mark were deriving from the Second-Order features group (7 out of 7 GLCM features – 100% of the features). A huge performance (in terms of rank repeatability over the different noGL) difference was noticed between the First-Order and Higher-Order features groups, in which just 3 out of 10 radiomic features (30% of the features) and 14 out of 25 radiomic features (56% of the features) were considered to have excellent repeatability, respectively. These results indicate that, even though the reliability of the vast majority of the radiomic features is not affected by the noGL, in some cases, their inter-ranking repeatability is being affected when different bin sizes are used. Leijenaar et al. [43] noticed the same results in a clinical study (instead of IQ phantoms, they used patients) comparing two different methodologies of setting bin size (FBS and FNB). Furthermore, they concluded that the interpretation of the radiomic features which do not show at least good rank reliability is highly dependent on the bin size – even if they seem promising, their prognostic and predictive value are heavily degraded. Hence, any conclusions deriving from these features (GLZLM features and GLRLM SRLGLE) at a given noGL cannot be directly compared when different bin sizes are applied, due to their poor reliability over the different noGL – their values are heavily affected by the total noGL.

The one-way ANOVA test was selected to determine whether the noGL affects the numeric value of the radiomic features. In the cases that a statistically significant difference ($p < 0.05$) was observed between the different noGL (64, 128, and 256 noGL) groups, an LSD test was incorporated to determine the specific grey level which causes the significant difference. The results of the one-way ANOVA test are presented in **Table 2.**, while LSD results are shown in **Table 3.** More specific, the values of all the features were heavily affected by the noGL, with just 5 out of 42 radiomic features (12% of the features in total) being robust –

these features were: First-Order features (Histogram Skewness, Kurtosis, and Excess Kurtosis), Second-Order features (GLCM Correlation), and Higher-Order features (NGLDM Coarseness). Most of the features showed a significant difference in all three pairwise comparisons of noGL. However, Second-Order feature (GLCM Contrast) and Higher-Order features (GLZLM SZLGLE and LZLGLE) showed no significant differences (at all ROIs) when comparing the measurements
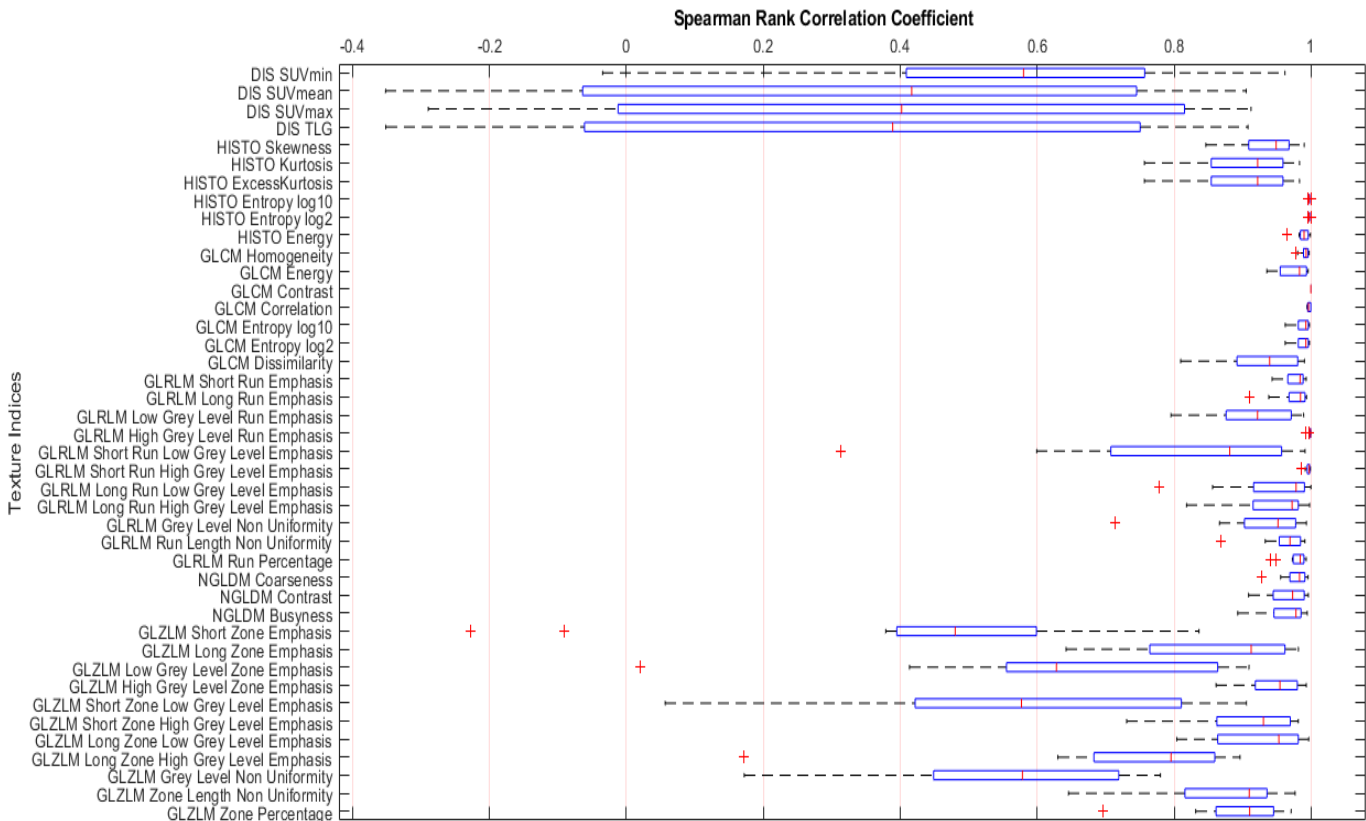
Figure 21. SPCC medians of the 2mm-diameter, 3mm-diameter, 4mm-diameter, and 5mm-diameter cylinders. SPCC values lower than 0.5 (but higher than 0) are indicating poor correlation (similarly values greater than -0.5 signify poor negative correlation), values ranging from 0.5 to 0.749 are indicating moderate correlation (similarly values ranging from -0.5-(-0.749) signify moderate negative correlation), values ranging from 0.75 to 0.899 are indicating good correlation (similarly values ranging from -0.75-(-0.899) signify good negative correlation), while values greater than 0.9 are indicating excellent correlation (similarly values lower than -0.9 signify excellent negative correlation). **Red cross:** outlier value, **Red vertical line:** median, **Blue Parallelogram:** range of values.

measurements of 128 noGL group to the measurements of 256 noGL group. These results indicate that the values of the most radiomic features are being affected by the noGL (as it has already been reported in the literature [43]) and just a couple of features (Histogram Skewness, Kurtosis, and Excess Kurtosis, GLCM Correlation, and NGLDM Coarseness) seems to be robust to the noGL applied. Moreover, by comparing the one-way ANOVA and LSD results from the ROIs of the cylinders to the ROI of the homogenous region, one can easily notice that the features which were robust to the noGL at cylinders were not robust at the homogenous region (HR) and vice versa – in the ROI of the HR, Higher-Order features (GLZLM GLNU and SZE, and NGLDM Contrast) were robust to the noGL. These results indicate that the ΔSUV (SUVmax-SUVmin) has a major role in the computation of the features – the ΔSUV of the ROI of the HR is approximately 0.40, while the ΔSUV of the ROIs of the cylinders is ranging from 2.10 to 3.10. Thus, the number of total bins existing in the ROI of the HR is way smaller than the number of bins in the ROIs of the cylinders (as seen in **Table 4.**), even though the applied bin size is the same which leads to different results. These results come in agreement with both Brancini et al. [80] and Van Velden et al [90] studies, which reported that a fixed bin size (FBS) should be preferred over a fixed number of bins (noGL) when doing texture analysis in patients. Another noteworthy conclusion which derives from this result is that different features seem to be robust at different numbers of total bins – Higher-Order features seem to be robust at a smaller number of bins (bigger bin sizes), while First-Order features and Second-Order features seem to be robust at greater numbers of bins. This observation could also mean that different texture indices may have similar prognostic or biological value in tumors with different SUV uptakes. Indeed, Foley et al. [86] conducted a clinical study and noticed that the conventional feature TLG did not have a prognostic value but it was linked with the overall survival in patients with esophageal cancer, while Feliciani et al. [87] reported that GLRLM-LRLGLE was linked with overall survival in patients with head and neck cancer. Cheng et al. [91] conducted a clinical study in patients with salivary gland adenocarcinoma and they were able to link overall survival with the Conventional feature TLG and Discretized intensity entropy.

Table 2. Results of the one-way ANOVA for the total of the acquisitions of the IQ phantom. Overview of the radiomic features which did not affect by the total noGL – these features showed no statistically significant difference (p>0.05) between the 64, 128, and 256 noGL groups.

Results of One-Way ANOVA - NEMA IQ Phantom Acquisitions				
2mm-diameter Cylinder	**3mm-diameter Cylinder**	**4mm-diameter Cylinder**	**5mm-diameter Cylinder**	**5mm-diameter Circle**
Histogram Skewness	Histogram Skewness	Histogram Skewness	Histogram Skewness	NGLDM Contrast
Histogram Kurtosis	Histogram Kurtosis	Histogram Kurtosis	Histogram Kurtosis	GLZLM SZE
Histogram Excess Kurtosis	Histogram Excess Kurtosis	Histogram Excess Kurtosis	Histogram Excess Kurtosis	GLZLM GLNU
GLCM Correlation	GLCM Correlation	GLCM Correlation	GLCM Correlation	
NGLDM Coarseness	NGLDM Coarseness	NGLDM Coarseness	NGLDM Coarseness	

COV and PD tests were calculated for a total of 51 radiomic features to determine the range of variance and the reproducibility of the features at all three noGL (64, 128, and 256 noGL). Results of the COV tests can be seen in **Figure 22.** Most of the features were found to be heterogeneous (COV>30%, have very low reproducibility at a given noGL), while only a limited number of features presented moderate variance or better. More specific, 31 out of 51 radiomic features (62% of the features) showed large variance in terms of COV, indicating that those features are heterogenous at all three noGL. Thus, they should not be considered liable and used for preclinical texture analysis because they are more prone to error (at the three bin sizes tested) – a conclusion supported by Galavis et al. [85] and Leijenaar et al. [89] who reported in their studies that radiomic features that present large variances cannot be used to quantify changes in the tumor's texture.

On the other hand, our results indicate that 20 out of 51 (38% of the features) radiomic features showed moderate variance or lower, with just 12 of 51 radiomic features presenting small variance or lower, these features were: First-Order features (5 out of 19 features – 42% of the features), Second-Order features (3 out of 7 features – 42% of the features), and Higher-Order features (4 out of 25 features – 16% of the features). Higher-Order features SZE and SRE, GLCM Entropy(s) and Correlation, and Histogram Entropy(s) have already been reported to be extremely robust features not only in terms of bin size but also when different reconstruction and acquisition parameters are applied, as mentioned by Galavis et al. [85]. These results are further supported by the data shown in **Figure 23.** (results of the PD test), indicating that texture indices presenting low variances are consistent at all noGL. Therefore, they can be further examined as potential biomarkers.

Table 3. Results of the LSD tests for the acquisitions of the IQ phantom. X symbol highlights a difference between the comparing groups which was not considered statistically significant.

		2mm-diameter Cylinder			3mm-diameter Cylinder			4mm-diameter Cylinder			5mm-diameter Cylinder			5mm-diameter Circle		
		64-128 noGL	64-256 noGL	128-256 noGL	64-128 noGL	64-256 noGL	128-256 noGL	64-128 noGL	64-256 noGL	128-256 noGL	64-128 noGL	64-256 noGL	128-256 noGL	64-128 noGL	64-256 noGL	128-256 noGL
Discretized	SUVmin	X			X			X			X					
	SUVmean															
	SUVmax															
	TLG															
Histogram	Skewness	X	X	X	X	X	X	X	X	X	X	X	X			
	Kurtosis	X	X	X	X	X	X	X	X	X	X	X	X			
	ExcessKurtosis	X	X	X	X	X	X	X	X	X	X	X	X			
	Entropy_lo10															
	Entropy_log2															
	Energy															
GLCM	Homogeneity															
	Energy			X			X									
	Contrast	X			X			X			X					
	Correlation	X	X	X	X	X	X	X	X	X	X	X	X			
	Entropy_log10															
	Entropy_log2															
	Dissimilarity													X		X
GLRLM	SRE															
	LRE			X												
	LGLRE						X			X			X			
	HGLRE	X														
	SRLGLE												X			X
	SRHGLE	X														
	LRLGLE			X	X		X			X			X			X
	LRHGLE	X														
	GLNU															
	RLNU															
	RP															
NGLDM	Coarseness	X	X	X	X	X	X	X	X	X	X		X			
	Contrast														X	
	Busyness						X			X						X
GLZLM	SZE	X												X	X	X
	LZE			X			X			X			X			
	LGLZE						X			X			X			
	HGLZE	X														
	SZLGLE			X			X			X			X			X
	SZHGLE	X			X			X						X		
	LZLGLE			X			X			X			X			X
	LZHGLE															
	GLNU						X			X				X		X
	ZLNU													X		X
	ZP															

Figure 22. Heatmaps of COVs of the radiomic features for each ROI at 64, 128, and 256 number of grey levels. COV values smaller or even to 5% indicate a very small variance, values ranging from 5% to 10% indicate a small variance, values ranging from 10% to 20% indicate a moderate variance, values ranging from 20% to 30% indicate a large variance, while values greater than 30% indicate that the given feature is heterogeneous.

Percent Differnce of 2mm-diameter Cylinder - 64 number of Grey Levels

Percent Differnce of 2mm-diameter Cylinder - 128 number of Grey Levels

Percent Differnce of 2mm-diameter Cylinder - 256 number of Grey Levels

23.1

Percent Differnce of 3mm-diameter Cylinder - 64 number of Grey Levels

Percent Differnce of 3mm-diameter Cylinder - 128 number of Grey Levels

Percent Differnce of 3mm-diameter Cylinder - 256 number of Grey Levels

23.2

Percent Differnce of 4mm-diameter Cylinder - 64 number of Grey Levels

Percent Differnce of 4mm-diameter Cylinder - 128 number of Grey Levels

Percent Differnce of 4mm-diameter Cylinder - 256 number of Grey Levels

23.3
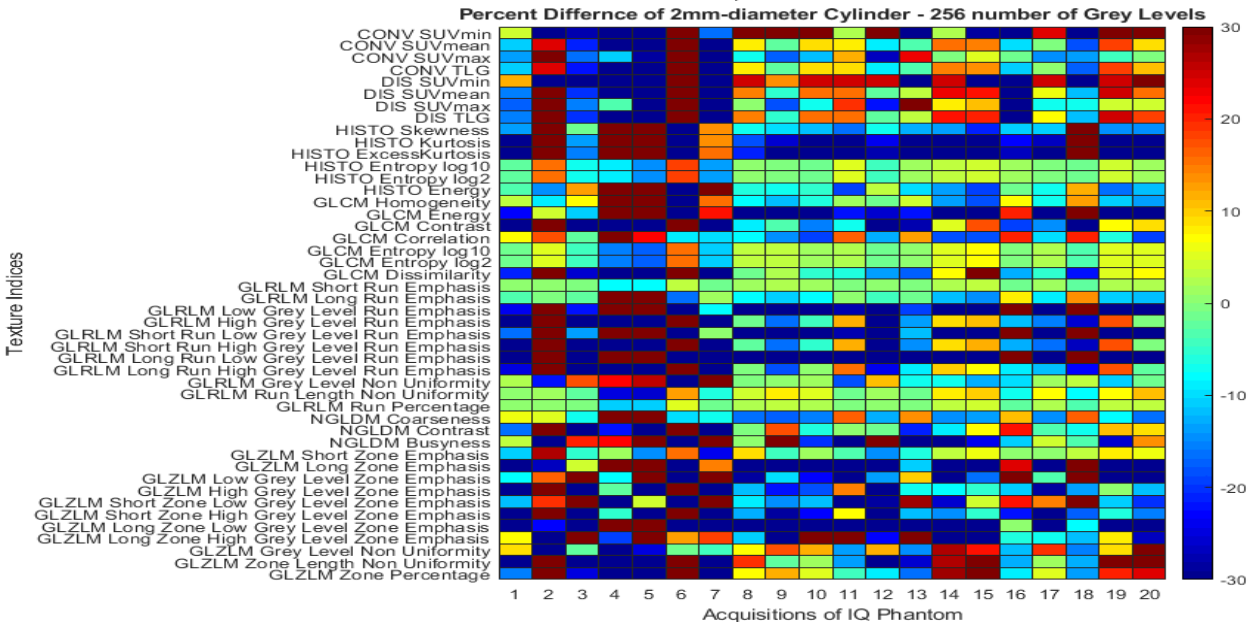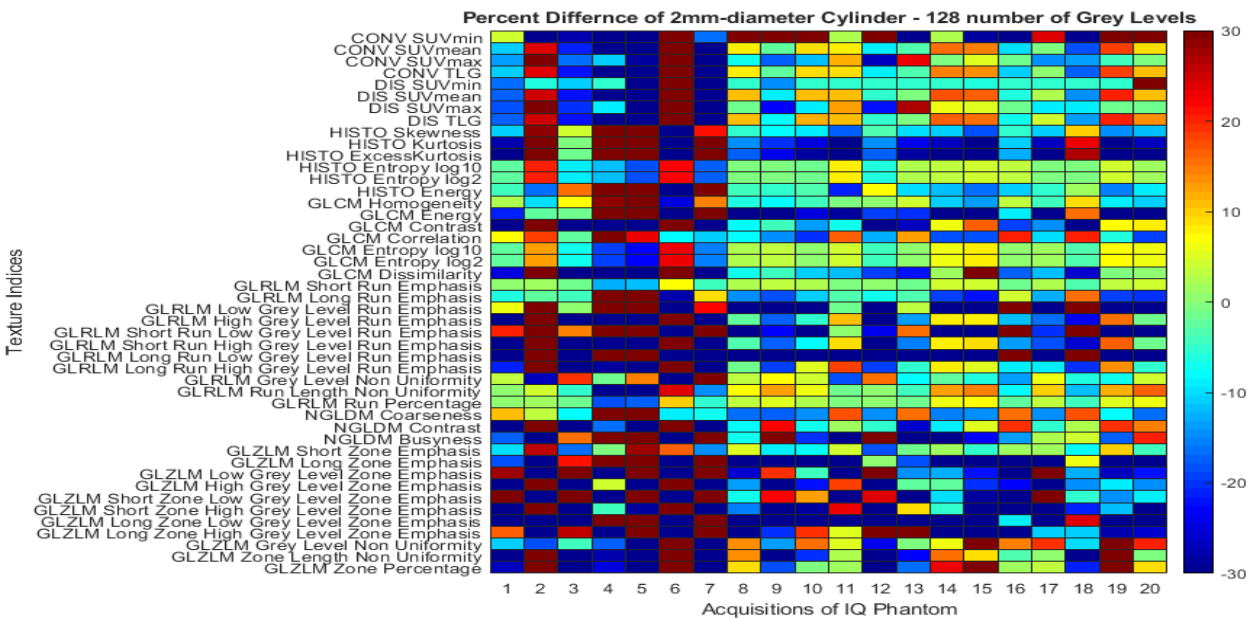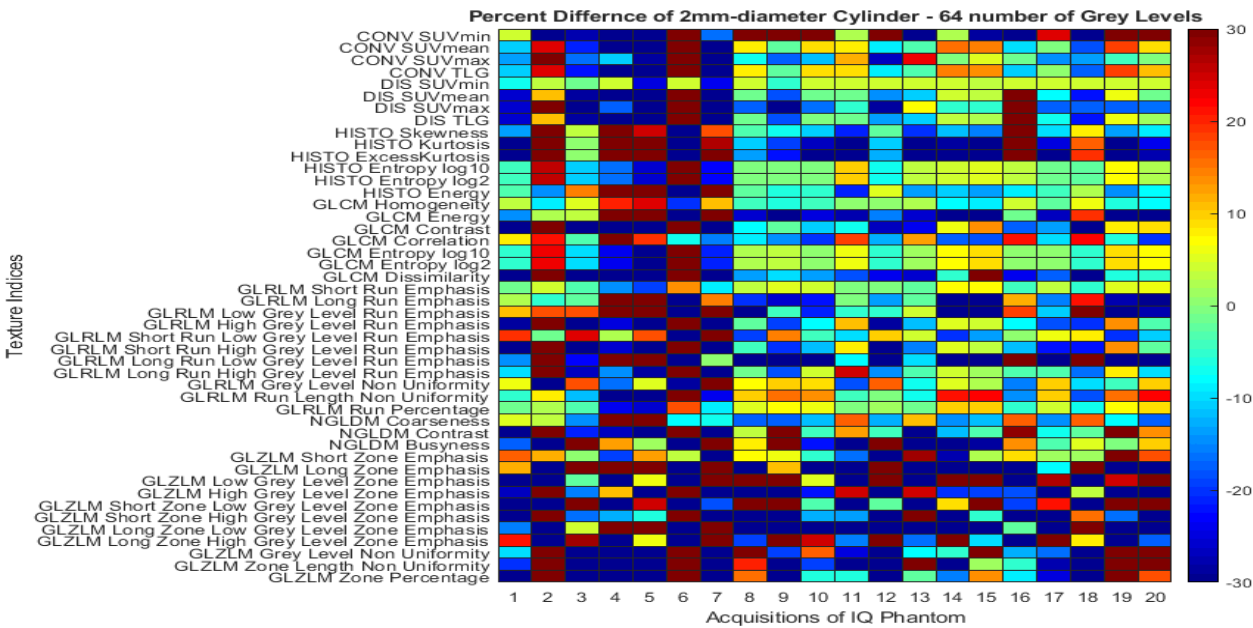
Percent Differnce of 5mm-diameter Cylinder - 64 number of Grey Levels

Percent Differnce of 5mm-diameter Cylinder - 128 number of Grey Levels

Percent Differnce of 5mm-diameter Cylinder - 256 number of Grey Levels
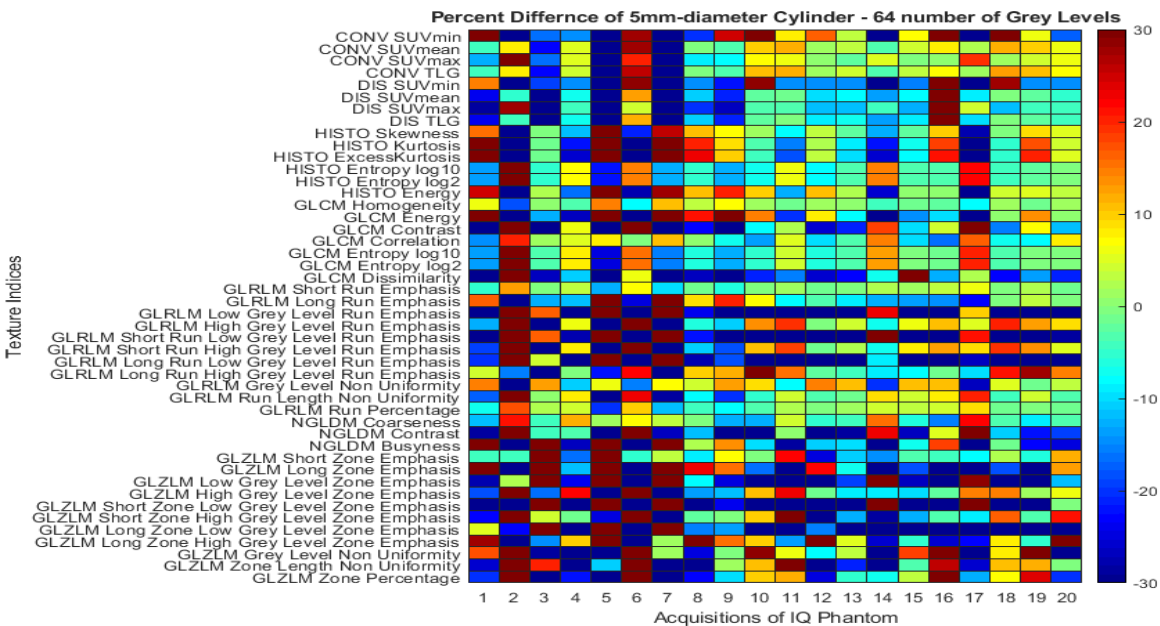
23.4

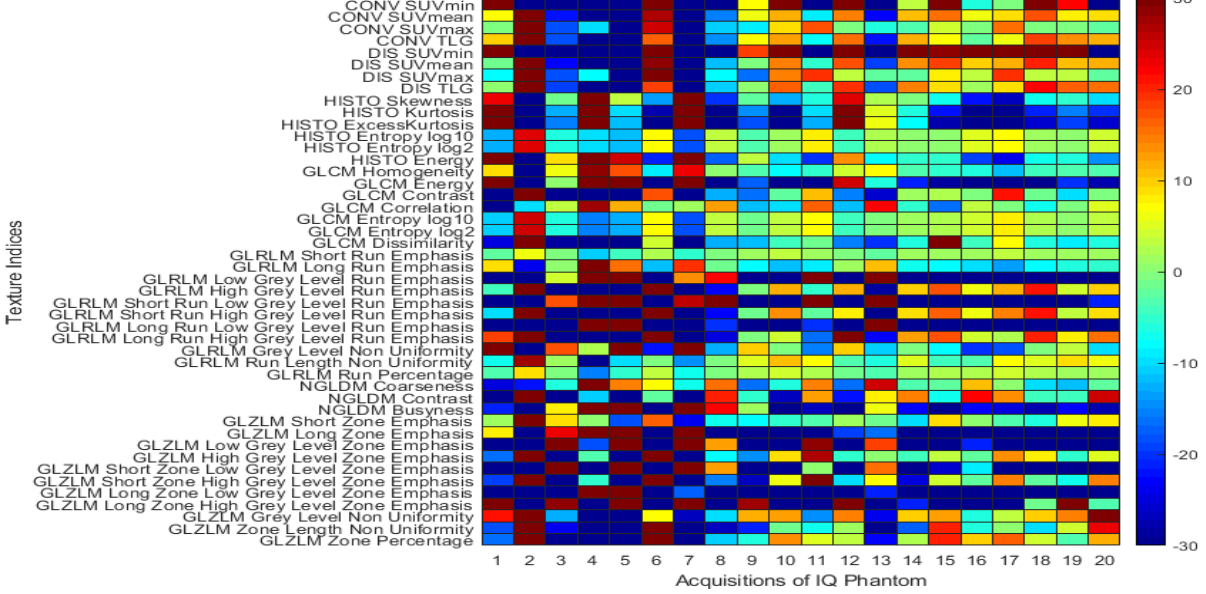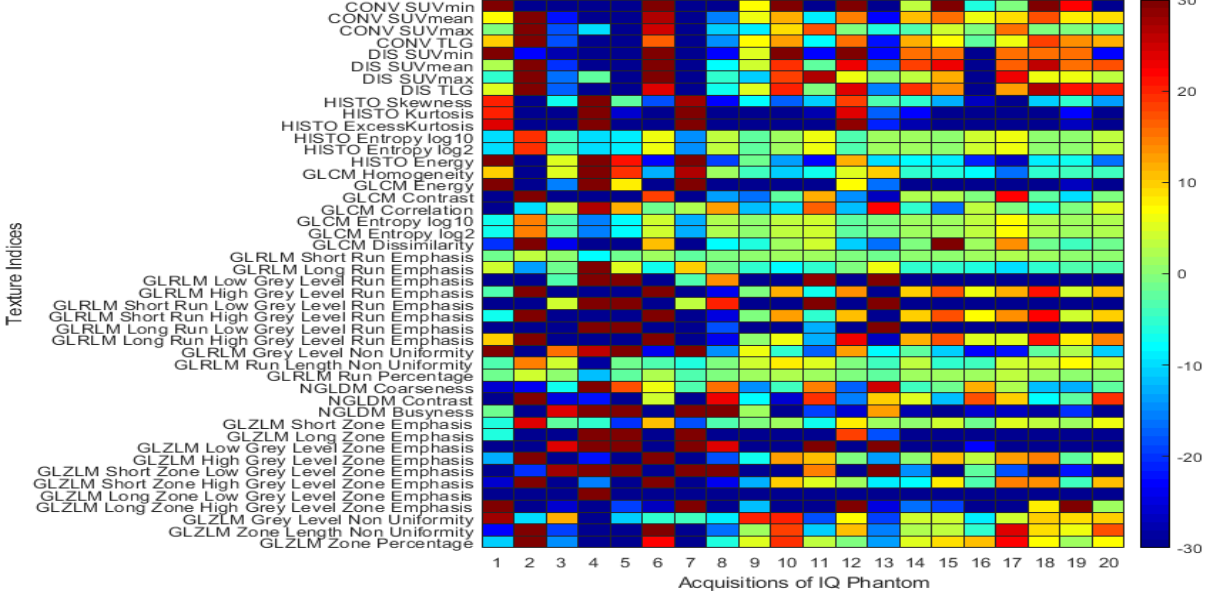Percent Differnce of 5mm-diameter Circle HR - 64 number of Grey Levels

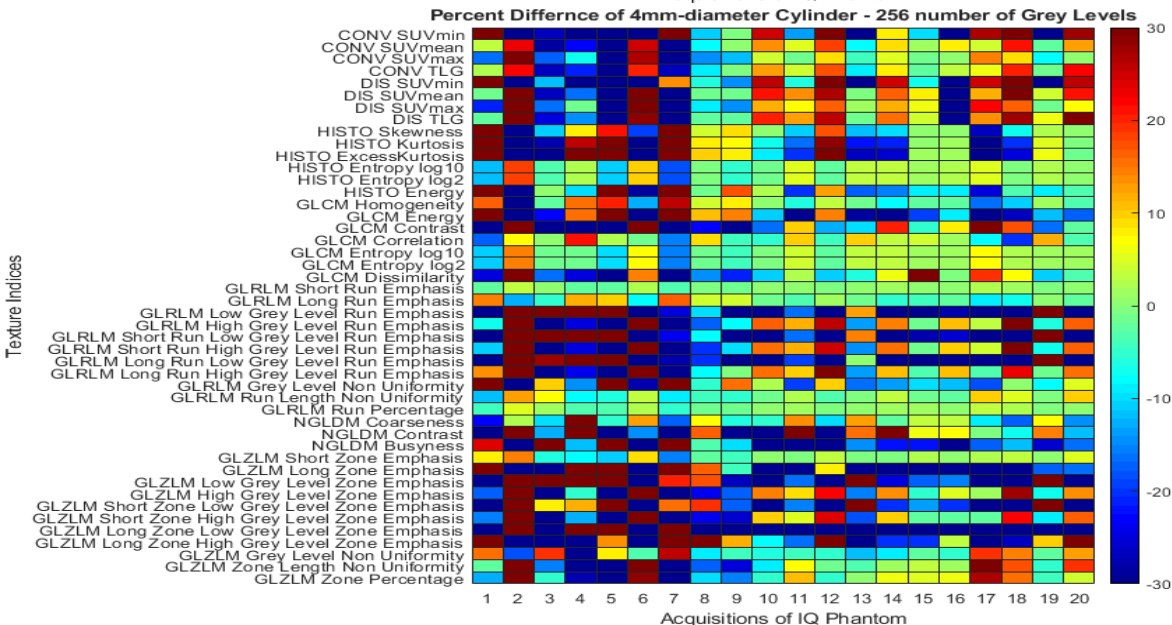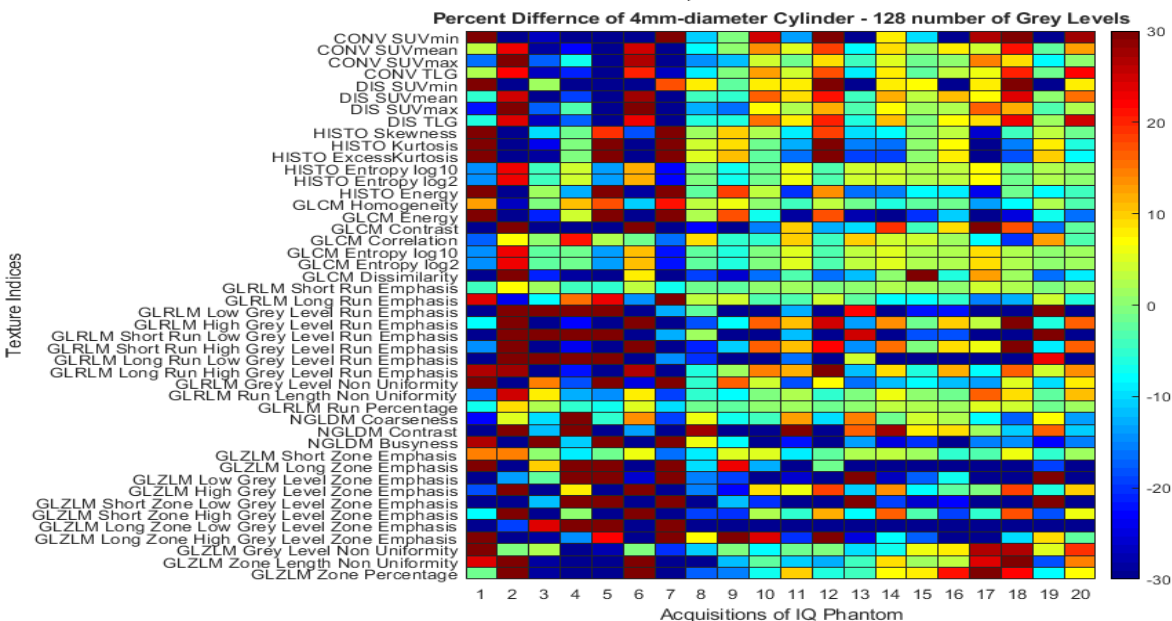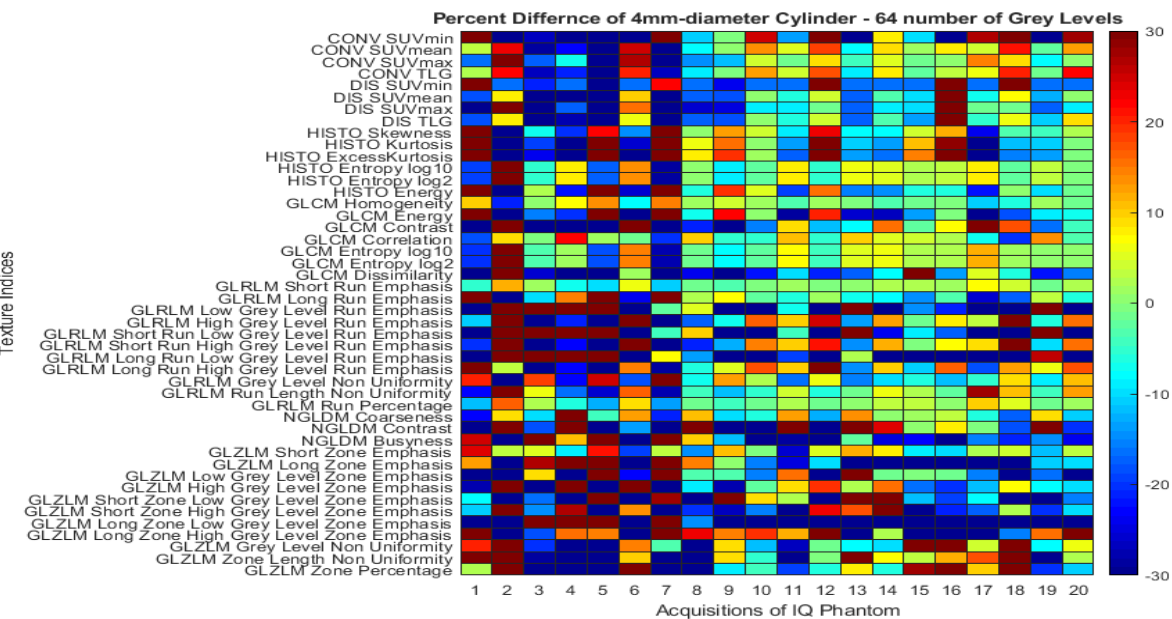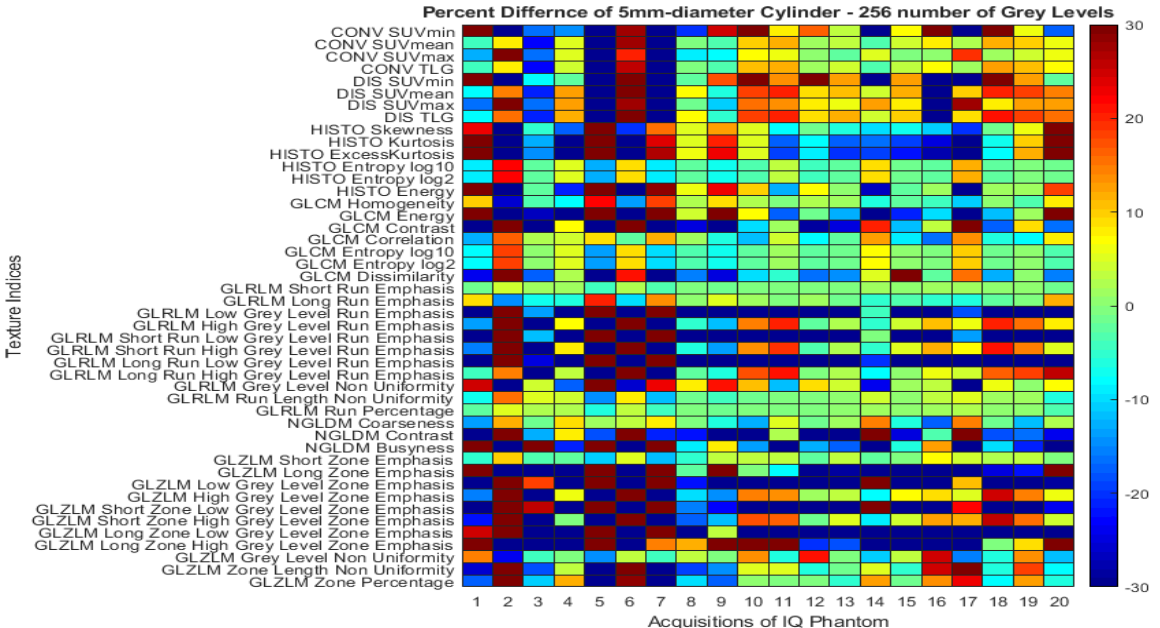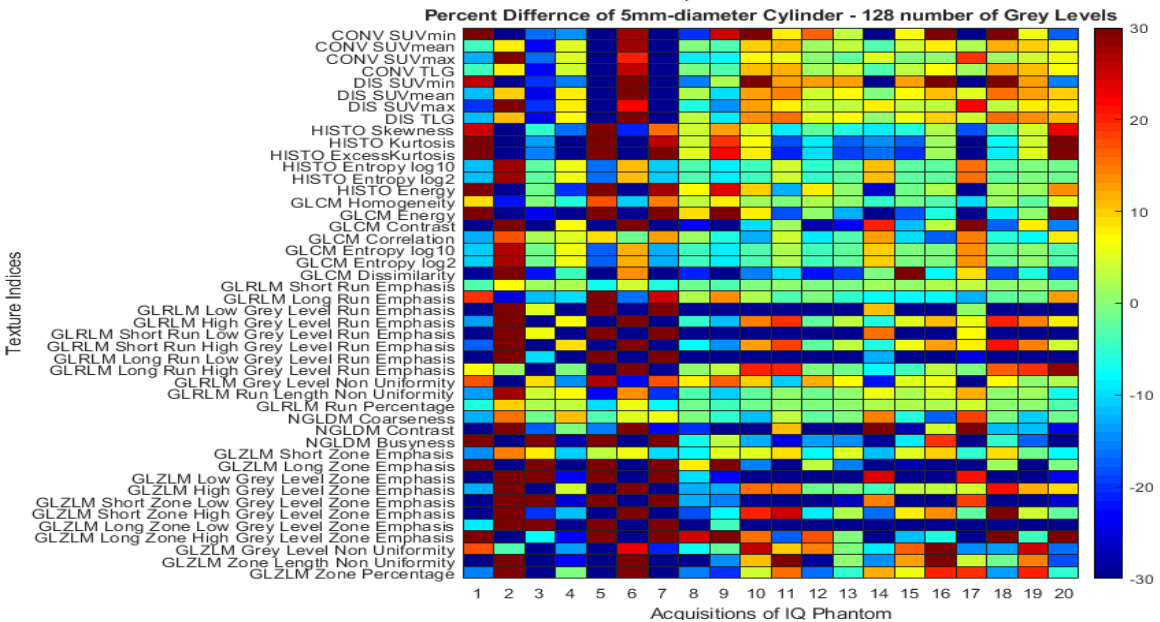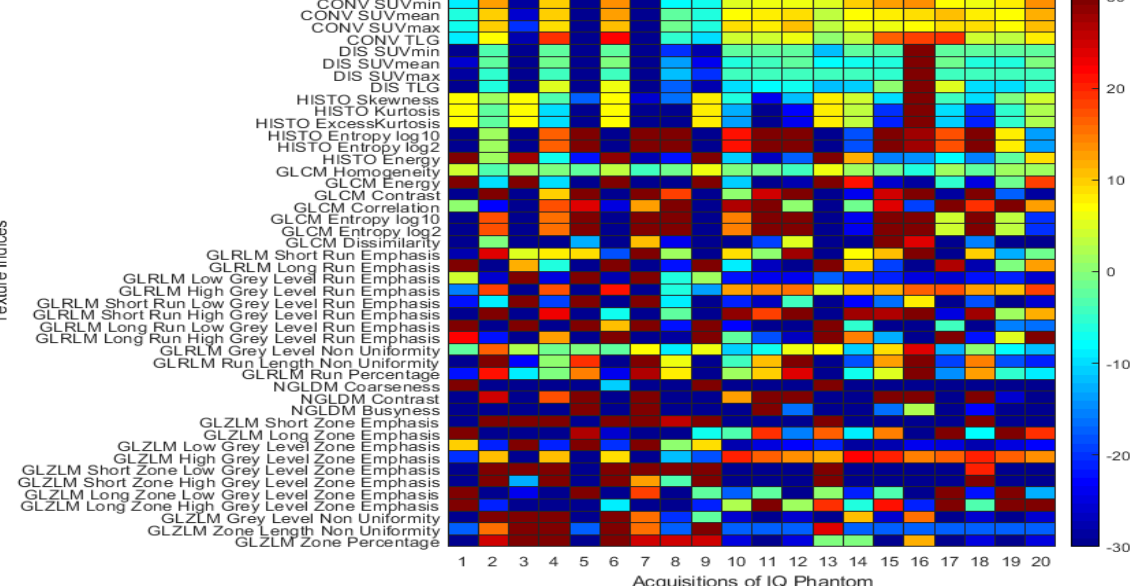
Percent Differnce of 5mm-diameter Circle HR - 128 number of Grey Levels
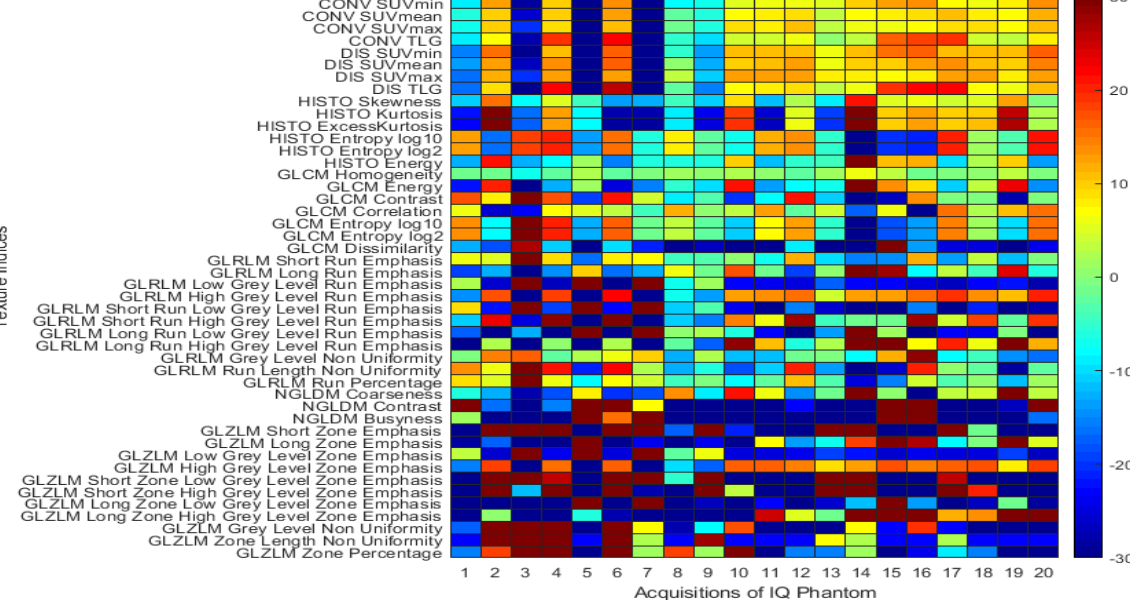
Percent Differnce of 5mm-diameter Circle HR - 256 number of Grey Levels
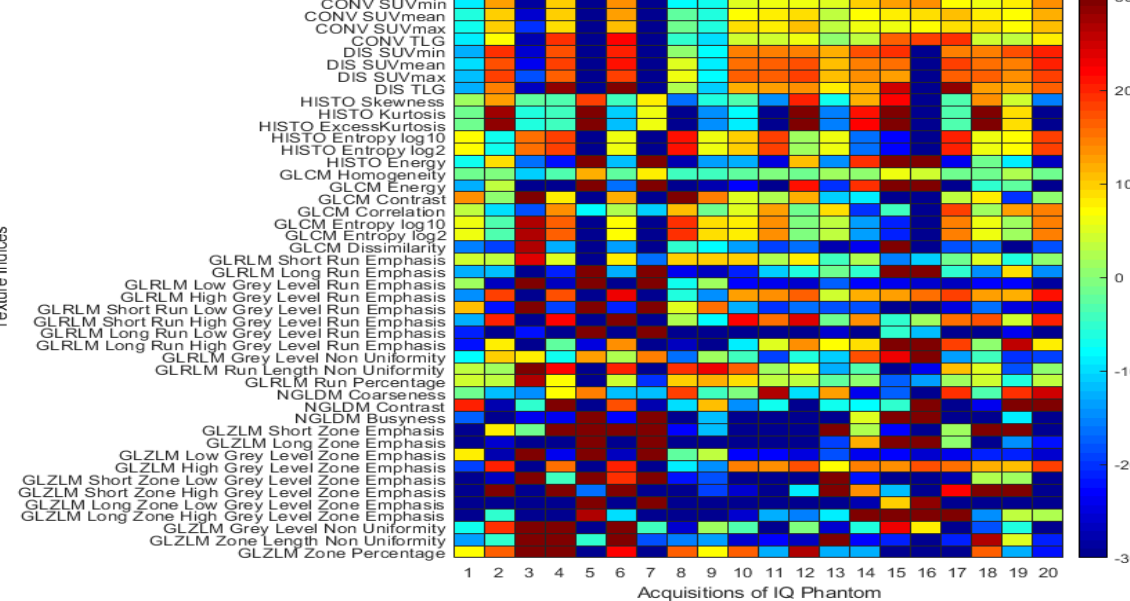
23.5

Figure 23. Heatmaps of PD for each ROI of the IQ phantom at 64, 128, and 256 noGL. PD

values lower than 5% indicate small variance, values ranging from 10% to 25% indicate

moderate variance, while values greater than 30% indicate large variance. **23.1** Shows the

results of PD tests of the ROI onto the 2mm-diameter Cylinder. **23.2** Shows the results of

PD tests of the ROI onto the 3mm-diameter Cylinder. **23.3** Shows the results of PD tests

of the ROI onto the 4mm-diameter Cylinder. **23.4** Shows the results of PD tests of the ROI

onto the 5mm-diameter Cylinder. **23.5** Shows the results of PD tests of the ROI onto the

5mm-diameter Circle in the homogenous region.

Table 4. Overview of size (in voxels), ΔSUV, and total number of bins (which exist in all three

bin sizes – 0.317, 0.157, and 0.078) of the ROIs.

Regions of Interest	Size (Voxels)	ΔSUV	Bins in 64 noGL	Bins in 128 noGL	Bins in 256 noGL
2mm-diameter Cylinder	700	2.114	7	14	27
3mm-diameter Cylinder	881	2.709	9	18	35
4mm-diameter Cylinder	1374	3.026	11	20	39
5mm-diameter Cylinder	1658	2.868	10	19	37
Homogenous Region	1057	0.388	2	3	5

Overall, the radiomic features which found to perform well at all 5 statistical tests

are Histogram Skewness, Kurtosis, and Excess Kurtosis, and Conventional feature

TLG (First-Order features), even though all of them showed moderate to large

variance, GLCM Correlation (Second-Order features), and NGLDM Coarseness

(Higher-Order features) – no GLRLM nor GLZLM features were able to perform

well at all 5 statistical tests, even SZE (which presented low variance (<10%) but

poor reliability in terms of ICC (>0.25)). Our data indicate that these radiomic features are consistent, reliable, and robust to all 3 noGL (64, 128, and 256) examined. Most of the features at ROIs with low SUV variation, they presented poor reliability as well as large variations as expected because of the extremely smaller number of total bins existing in these ROIs which led to more extreme and inconsistent values – a conclusion that comes in agreement with Branchini et al. [80] and Van Vendel et al. [90] studies, who both reported that at higher bin sizes (low noGL) only a limited number of radiomic features presented increased robustness. The data shown in this study, suggests that direct comparison of radiomic features should be performed only when the examined ROIs have similar sizes (equal number of voxels) and can be split into an equal number of bins (either have similar ΔSUVs or incorporation of different bin sizes to end up with a similar number of bins at each ROI). However, according to the basic theory of Histogram in Statistics [93] different bin sizes reflect different information – radiomic features may be similarly affected. Hence, further studies are required to assess whether different bin sizes can reveal or conceal data, and which number(s) of bins is(are) the optimal for the texture analysis if any.

3.2 Animals Study

One-way ANOVA test was also incorporated to determine whether the noGL affects the numeric values of the radiomic features in animals (mice), as well as examine which feature values are affected by the administration of treatment. Since, no animals received treatment till the acquisition of the 10th day, Treated and Non-Treated mice on 7 and 10 days were merged to a single group for each set of acquisitions (7 days' group and 10 days' group). Comparisons between the different noGL was performed on the mice of the 7 days' group, 10 days' group, Treated group (4 days after the administration of treatment) on 14 days, and Non-Treated group (no treatment received) on 14 days, to determine the impact of the noGL on the values of the features. In the cases that a statistically significant difference ($p<0.05$) was observed between the different noGL (64, 128, and 256

noGL), an LSD test was incorporated to determine the specific grey level which causes the significant difference. Moreover, straight comparisons of the acquisitions of both Treated and Non-Treated mice between day 10 and 14 at each noGL (64, 128, and 256) was performed to determine which features were affected by the administration of the treatment (showed statistical difference ($p<0.05$) only in the comparison of the Treated group on 10 and 14 days) and which features could be an indicator of the progress of the disease (showed statistical difference ($p<0.05$) only in the comparison of the Non-Treated group on 10 and 14 days).

Table 5. Results of the one-way ANOVA for the total of the acquisitions of the IQ phantom. Overview of the radiomic features which did not affect by the total noGL – these features showed no statistically significant difference ($p>0.05$) between the 64, 128, and 256 noGL groups.

Results of One-Way ANOVA - Mice			
7 days	10 days	14 days Treated	14 days Non-Treated
Histogram Skewness	Histogram Skewness	DIS SUVmin	DIS SUVmin
Histogram Kurtosis Histogram Excess Kurtosis	Histogram Kurtosis Histogram Excess Kurtosis	Histogram Skewness	Histogram Skewness
		Histogram Kurtosis Histogram Excess Kurtosis	Histogram Kurtosis Histogram Excess Kurtosis
GLCM Correlation	GLCM Correlation	GLCM Correlation	GLCM Correlation
NGLDM Coarseness	GLRLM RLNU	GLRLM LRHGLE	GLRLM SRE
	NGLDM Coarseness	GLRLM GLNU	GLRLM SRHGLE
	GLZLM LZLGLE	GLRLM RLNU	GLZLM LZLGLE
		NGLDM Coarseness	
		GLZLM LZE	
		GLZLM LZLGLE	
		GLZLM LZHGLE	
		GLZLM ZLNU	

As seen in **Table 5.**, similar to the results of the one-way ANOVA for the acquisitions of the IQ phantom, all the features showed low-to-no robustness regarding the noGL used. Only features which were found to be robust at all 4 comparisons (on the mice of the 7 days' group, 10 days' group, Treated group on 14 days, and Non-Treated group on 14 days), were First-Order features (Histogram Skewness, Kurtosis, and Excess Kurtosis), Second-Order features (GLCM Correlation), and Higher-Order features (NGLDM Coarseness). Also, GLRM RLNU showed great robustness at the mice of 7 and 10 days' group, as well as at the mice of the Treated group on 14 days, but showed no robustness at the mice of the Non-Treated group on 14 days. Furthermore, a couple of GLZLM features (LZE, LZLGLE, LZHGLE, ZLNU) were found to be robust at the mice of the Treated group on 14 days (but not in any other case). An LSD test was incorporated in the cases where a statistically significant difference was noticed to determine between which groups of noGL vary significantly (as seen in **Table 6.**). For the most features, no statistically significant difference was noticed between the comparisons of 64 and 128 noGL groups and 128 and 256 noGL groups. However, nearly all the comparisons between the features of the 64 and 256 noGL groups found to be statistically significant. These results come in agreement with the findings of the IQ phantom study. Also, the same features (Histogram Skewness, Kurtosis, and Excess Kurtosis, GLCM Correlation, and NGLDM Coarseness) were found to be robust in all cases. Additionally, the ΔSUV of the mice in any case was greater than 2.0 which further supports the previous conclusion that First-Order and Second-Order features tend to be more robust at larger number of bins (higher SUV variations). Higher-Order features (GLZLM LZE, LZLGLE, LZHGLE, ZLNU) were found to be robust only at the Treated mice group on 14 days and not in any other group.

Regarding the results of one-way ANOVA between the mice of the Treated and Non-Treated group on 10 and 14 days (as seen in **Table 7.**), a limited amount of features showed statistically significant difference solely on the mice of the Treated group – these features were: First-Order features (Histogram Skewness, both Entropy log2 and log10, and Energy), Second-Order features (GLCM

Homogeneity, Energy, Correlation, and both Entropy log2 and log10), and Higher-Order features (GLRLM SRE, LRE, and RP, and NGLDM Busyness). Conventional TLG was the only feature which showed statistical significance between day 10 and 14 in the Non-Treated group and not in the corresponding comparison of the Treated group. These results come in agreement with the Scalco et al. [37] review, which concluded that the features reflecting intra-tumoral heterogeneity provide more valuable information about overall survival and treatment response, as well as may have some correlation with cell cycling pathways. This claim is further supported by the evidence shown in other works [43,80,88] in which radiomic features measuring heterogeneity have been reported to have a prognostic value and being indices of overall survival.

Table 6. Results of the LSD tests for the acquisitions of the mice over the course of the 14 days. X symbol highlights a difference between the comparing groups which was not considered statistically significant.

		7 days			10 days			14 days Treated			14 days Non-Treated		
		64-128 noGL	64-256 noGL	128-256 noGL	64-128 noGL	64-256 noGL	128-256 noGL	64-128 noGL	64-256 noGL	128-256 noGL	64-128 noGL	64-256 noGL	128-256 noGL
Discretized	SUVmin	X			X		X	X	X	X	X	X	X
	SUVmean							X					
	SUVmax							X			X		
	TLG	X			X			X		X	X		
Histogram	Skewness	X	X	X	X	X	X	X	X	X	X	X	X
	Kurtosis	X	X	X	X	X	X	X	X	X	X	X	X
	ExcessKurtosis	X	X	X	X	X	X	X	X	X	X	X	X
	Entropy_lo10												
	Entropy_log2												
	Energy						X						
GLCM	Homogeneity												
	Energy			X			X			X			X
	Contrast	X			X			X			X		
	Correlation	X	X	X	X	X	X	X	X	X	X	X	X
	Entropy_log10												
	Entropy_log2												
	Dissimilarity							X			X		
GLRLM	SRE						X						X
	LRE			X			X			X	X	X	X
	LGLRE			X			X			X			X
	HGLRE	X			X			X			X		
	SRLGLE			X			X			X			X
	SRHGLE	X			X			X			X		
	LRLGLE			X			X			X	X	X	X
	LRHGLE	X			X			X		X	X		
	GLNU	X		X			X	X	X	X			
	RLNU	X		X	X	X	X	X	X	X			
	RP						X						X
NGLDM	Coarseness	X	X	X	X	X	X	X	X	X	X		X
	Contrast	X						X					
	Busyness			X			X			X			X
GLZLM	SZE												
	LZE			X			X	X	X	X			X
	LGLZE			X			X			X			X
	HGLZE	X			X			X			X		
	SZLGLE			X			X			X			X
	SZHGLE	X			X			X			X		
	LZLGLE			X	X	X	X	X	X	X	X	X	X
	LZHGLE			X			X	X	X	X			X
	GLNU	X					X	X					
	ZLNU	X			X			X	X	X	X		
	ZP	X						X			X		

To examine the reproducibility and liability of the features in terms of robustness over the different noGL applied, a COV test was incorporated (as seen in **Figure 24.**). The results of the mice were similar to the results of the IQ phantom study. Even though some indices showed similar repeatability in both studies, other indices did not. More specific, all the GLZLM and GLRLM features showed extremely low variance on the mice of the Treated group on day 14, while on the rest of the groups the same features were considered heterogeneous – a result which was noticed in the IQ phantom study as well – indicating that the administration of treatment may have influenced the repeatability of the features. Other features such as GLCM Correlation and GLRLM SRE and RP have similar performances in both studies. More specific, the features with the lowest variances in all cases were First-Order features (Histogram Skewness (<10%) and both Entropy lo2 and log10 (<5% at mice of Non-Treated group on 14 days, rest of the groups <20%)), Second-Order features (GLCM Correlation (<5%), Homogeneity (<20%) and both Entropy log2 and log10 (<20%)), and Higher-Order features (GLRLM SRE (<10%) and RP (<10%)). It is mention-worthy that both Higher-Order features (all GLRLM and GLZLM features) showed a COV lower than 0.1% only at the mice of the Treated group on 14 days, while First-Order features (Conventional TLG and SUVmax) showed moderate to high variance only at the mice of Non-Treated on 14 days – in general, these features showed extremely high variances.

 Only a limited number of features found to potentially have a prognostic value and could be further investigated in future studies are GLCM-Correlation and Histogram-Skewness, as they were the only that showed excellent reproducibility (at every noGL applied), increased robustness over the different noGL applied, and their numeric values were heavily affected by the administration of treatment (mice of Non-Treated group showed no statistically significant difference in the comparison of the days 10-14). This is not the first time that GLCM-Correlation was proposed as a reliable and a good candidate in PET imaging, as Leijenaar et al. [43] noticed that it was the only texture feature that its ranking was consistent in between the patients, while it was robust both to SUV discretization methods and

criteria – which comes in agreement with our results. Furthermore, in a recent study Najako et al. [88] reported that GLCM-Entropy in combination with two other texture features (GLCM-Dissimilarity and CONVENTIONAL-SUVmean) can be used to distinguish patients with malignant tumors from those with benign tumors, in the respiratory system. Another clinical study conducted by Foley et al. [80], reported Histogram Energy and Kurtosis as well as the Conventional feature TLG did not have a prognostic value but they were linked with overall survival in patients with esophageal cancer – which comes in agreement (partially) with our results (as HISTO-Energy was found to have excellent SPCC and ICC scores in the IQ phantom study, but showed a moderate to a large variance in the COV and PD tests and a high dependency on noGL on both studies, however, it did show a statistically significant difference between days 10-14 at the mice of the Treated group but not in the mice of the Non-Treated group). Feliciani et al. [90] carried out a clinical study on head and neck cancer, noticing that GLRLM-LRLGLE did have a prognostic value and was closely related to overall survival which does not agree with our results, as GLRLM-LRLGLE showed excellent reliability and rank reliability and dependency on noGL (at 64 and 128 noGL), but large variances (both in terms of COV and PD tests) on both studies. However, this difference is not surprising and further supports that different indices may describe each cancer type.

Table 7. Results of the one-way ANOVA between the day 10 and 14 for the mice of both the Treated and Non-Treated group. **7.1** Overview of the features which presented statistical significance (p<0.05) between day 10 and 14 on the Treated group. **7.2** Overview of the features which presented statistical significance (p<0.05) between day 10 and 14 on the Non-Treated group.

Results of One-Way ANOVA - Treated Mice on days 10 & 14		
64 noGL	128 noGL	256 noGL
Histogram Skewness	Histogram Skewness	Histogram Skewness
Histogram Entropy log10	Histogram Entropy log10	
Histogram Entropy log2	Histogram Entropy log2	
Histogram Energy	Histogram Energy	
GLCM Homogeneity	GLCM Homogeneity	
GLCM Energy	GLCM Energy	
GLCM Contrast	GLCM Contrast	
GLCM Entropy log10	GLCM Entropy log10	
GLCM Entropy log2	GLCM Entropy log2	
GLCM Dissimilarity	GLCM Dissimilarity	
GLRLM SRE	GLRLM SRE	
GLRLM LRE	GLRLM LRE	
GLRLM LRLGLE	GLRLM GLNU	
GLRLM RP	GLRLM RP	
NGLDM Contrast	NGLDM Contrast	
NGLDM Busyness	NGLDM Busyness	
	GLZLM ZP	

7.1

Results of One-Way ANOVA - Non-Treated Mice on days 10 & 14		
64 noGL	128 noGL	256 noGL
Conventional TLG	Conventional TLG	Conventional TLG
DIS TLG	DIS TLG	DIS TLG
Shape Volume	Shape Volume	Shape Volume
Shape Surface	Shape Surface	Shape Surface
GLCM Homogeneity	GLCM Contrast	GLCM Contrast
GLCM Contrast	GLCM Dissimilarity	GLCM Dissimilarity
GLCM Entropy log10	GLRLM GLNU	GLRLM GLNU
GLCM Entropy log2	NGLDM Contrast	NGLDM Contrast
GLCM Dissimilarity		GLZLM ZP
GLRLM LRHGLE		
GLRLM GLNU		
NGLDM Contrast		
GLZLM LZHGLE		

7.2

Figure 27. Heatmaps of COVs of the radiomic features for the mice of all 4 groups over the course of the 14 days, at 64, 128, and 256 number of grey levels. COV values smaller or even to 5% indicate a very small variance, values ranging from 5% to 10% indicate a small variance, values ranging from 10% to 20% indicate a moderate variance, values ranging from 20% to 30% indicate a large variance, while values greater than 30% indicate that the given feature is heterogeneous.

This study has a couple of limitations, one of these is the inter-scanner variability as a single PET/CT scanner was incorporated (Mediso nanoScan®PC (PET8/2)). It is already well-proven that texture indices vary between different scanners, as being reported by Ger et al. [92]. However, the reconstruction and acquisition parameters were standardized and selected after a study conducted by Gaitanis et al. [76] on the specific scanner. Another limitation of the study is the fact that all animals used in the experiments were male – thus, the impact of the gender wasn't investigated. Also, the sample size of treated (6) and untreated rats (4) was limited. Since standard protocols for the evaluation of the radiomic features do not exist, this could be considered another limitation, even though the study was planned according to previous experience found in academic literature [41,43,80,85,86,87,89]. Other issues with this study – when it comes to the preliminary phase of the study – is the limited intra-observer variability, as the assessment of the tumors' area both before- and post-treatment were done by a single rater and not three, however, the cancer type investigated was melanoma which is characterize by clearly visible tumors. The injected activity of 18F-FDG can be considered as a limitation, as just a certain activity (3.7 MBq for the acquisition of the IQ phantom) was incorporated – different activities may have a strong influence on the computation of the radiomic features.

4. Conclusion

The purpose of this work was the accurate identification and validation of texture indices deriving from small-animal PET image analysis that could serve as potential biomarkers, a process characterized by the lack of standardized methodology for both the acquisition parameters of the image and the evaluation protocols of the generated texture indices. We focused on the impact of SUV discretization using three total fixed number of bins (64, 128, and 256 number of grey levels). The current research was split into two phases. In the first phase, texture analysis was performed onto 20 unique acquisition of a NEMA IQ phantom, while in the second phase a preliminary study was conducted on animals (mice were split into two groups: Treated and Non-Treated). Our analysis shown that the vast majority of the radiomic features is heavily affected by the employed number of grey levels. Only a limited number of features (Histogram Skewness, Excess Kurtosis and Kurtosis, GLCM Correlation, and NGLDM Coarseness) showed excellent reliability, reproducibility, liability, and robustness over the different number of grey levels applied in the IQ phantom study, while just Histogram-Skewness and GLCM-Correlation were able to achieve statistically significant results between the Treated and Non-Treated group of the mice (while being robust). This study shows that the manner of the SUV discretization has a crucial effect on the computation of the texture features and suggests a unique way of validation of the texture features that could serve as potential biomarkers – underlining the need for a standardized methodology in tumor texture analysis.

This is the very first study which claims that different texture indices may be linked with similar biological characteristics (Higher-Order feature tend to be robust at highly homogenous tumors (low SUV variations), while certain First-Order and Second-Order features tend to be robust at heterogeneous tumors (high SUV variations)) due to the different properties of different cancer types – different SUV uptakes lead to different SUV variations in the tumor regions, which may reflect differences in the biochemical or genomic profile of the tumor (potential linking between the information derived from statistical analysis of the

distribution of the radiopharmaceutical). Further research is needed for the evaluation of this suggestion. If this proved true, then radiopharmaceuticals which distribution is determined by different biological/physiological factors may provide additional information through the same texture indices or/and different texture indices may prove robust and reliable.

5. Future Perspectives

 Further work can be performed to investigate the effect of SUV discretization (bin size – number of grey levels) on the computation of the texture features and the identification of an optimal bin size which reveals more information than it conceals and whether it can be applicable on all the cancer types or a unique bin size should be selected for every tumor, based on tumor's unique biochemical and genomic properties. More extensive and long-term preclinical studies must be conducted to observe how radiomic features tend to evolve over time (the studies should feature both healthy and non-healthy candidates), while other studies should focus more on new and more complex ways of validation of the texture features (which could serve as potential biomarkers) – every time an issue is addressed, a ton of others come up. Finally, a standardized methodology should be proposed to limit the discrepancy between the different works and have reliable and comparable results. Even though, the field of Radiomics is – at very least – questionable and kind of arbitrary, there is a strong potential to have biomarkers deriving from texture analysis of the medical images that could provide information not only for the current or the near-future conditions of the patient but for more long-term health issues (e.g. future cardiac or respiratory issues, organ failure expectancy, life expectancy). After all, the world is ruled by quantum mechanics, the information (in all of its forms) is quantized – why not being present in the images too?

6. References

1. Guide to Medical Image Analysis – ISBN: 978-1-4471-2751-2

2. Carovac A, Smajlovic F, Junuzovic D. – Application of ultrasound in medicine. - Acta Inform Med. 2011;19(3):168-171. doi:10.5455/aim.2011.19.168-171

3. Imaginis

4. US Department of Energy – Molecular Nuclear Medicine Legacy

5. Berger A. – Positron Emission Tomography – BMJ. 2003 Jun 28; 326(7404): 1449. doi:10.1136/bmj.326.7404.1449

6. Belhocine, T., Spaepen, K., Dusart, M., Castaigne, C., Muylle, K., Bourgeois, P.,Flamen, P. (2006). 18FDG PET in oncology: The best and the worst (Review). International Journal of Oncology. doi:10.3892/ijo.28.5.1249

7. Castelli, J., De Bari, B., Depeursinge, A., Simon, A., Devillers, A., Roman Jimenez, G., Prior, M., Ozsahin, R., de Crevoisier Bourhis, J. (2016). Overview of the predictive value of quantitative 18 FDG PET in head and neck cancer treated with chemoradiotherapy. Critical Reviews in Oncology/Hematology, 108, 40–51. doi:10.1016/j.critrevonc.2016.10.009

8. Parkin, D.M., Bray, F., Ferlay, J., Pisani, P., 2005. Global cancer statistics, 2002. CA Cancer J Clin 55(2), 74-108

9. Yu, W., Cai, X.-W., Liu, Q., Zhu, Z.-F., Feng, W., Zhang, Q., Zhang, Y., Yao, Z., Fu, X.-L. (2015). Safety of dose escalation by simultaneous integrated boosting radiation dose within the primary tumor guided by 18FDG-PET/CT for esophageal cancer. Radiotherapy and Oncology, 114(2), 195–200. doi:10.1016/j.radonc.2014.12.007

10. Esophageal Cancer Factsheet. Global Cancer Observatory. Retrieved 8 November 2019

11. Sun, X.-Y., Chen, T.-X., Chang, C., Teng, H.-H., Xie, C., Ruan, M.-M., Lei, B., Liu, L., Wang, L.,Yang, L., Xie, W.-H. (2020). SUVmax of 18FDG PET/CT Predicts Histological Grade of Lung Adenocarcinoma. Academic Radiology. doi:10.1016/j.acra.2020.01.030

12. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018), Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68: 394-424. doi:10.3322/caac.21492

13. Nakamura, H., Saji, H., Shinmyo, T., Tagaya, R., Kurimoto, N., Koizumi, H., & Takagi, M. (2015). Close association of IASLC/ATS/ERS lung adenocarcinoma subtypes with glucose-uptake in positron emission tomography. Lung Cancer, 87(1), 28–33. doi:10.1016/j.lungcan.2014.11.010

14. Groheux, D., Martineau, A., Teixeira, L., Espié, M., de Cremoux, P., Bertheau, P., Merlet, P., Lemarignier, C. (2017). 18FDG-PET/CT for predicting the outcome in ER+/HER2- breast cancer patients: comparison of clinicopathological parameters and PET image-derived indices including tumor texture analysis. Breast Cancer Research, 19(1). doi:10.1186/s13058-016-0793-2

15. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D., Piñeros, M., Znaor, A. and Bray, F. (2019), Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int. J. Cancer, 144: 1941-1953. doi:10.1002/ijc.31937

16. Pace, L., Nicolai, E., Basso, L., Garbino, N., Soricelli, A., & Salvatore, M. (2020). Brown Adipose Tissue in Breast Cancer Evaluated by [18F] FDG-PET/CT. Molecular Imaging and Biology. doi:10.1007/s11307-020-01482-z

17. Vaidyanathan S, Patel CN, Scarsbrook AF, Chowdhury FU. FDG PET/CT in infection and inflammation--current and emerging clinical applications. Clin Radiol. 2015;70(7):787-800. doi:10.1016/j.crad.2015.03.010

18. Warburg O. Metabolism of tumours. Biochem Z 1923;142:317e33.

19. Kubota R, Yamada S, Kubota K, et al. Intratumoral distribution of fluorine-18-fluorodeoxyglucose in vivo: high accumulation in macrophages and granulation tissues studied by microautoradiography. JNucl Med 1992;33(11):1972e80.

20. APA PETERSDORF, ROBERT G.; BEESON, PAUL B. FEVER OF UNEXPLAINED ORIGIN: REPORT ON 100 CASES, Medicine: February 1961 - Volume 40 - Issue 1 - p 1-30.

21. Kouijzer, I. J. E., Bleeker-Rovers, C. P., & Oyen, W. J. G. (2013). FDG-PET in Fever of Unknown Origin. Seminars in Nuclear Medicine, 43(5), 333–339. doi:10.1053/j.semnuclmed.2013.04.005

22. Van der Bruggen, W., Bleeker-Rovers, C. P., Boerman, O. C., Gotthardt, M., & Oyen, W. J. G. (2010). PET and SPECT in Osteomyelitis and Prosthetic Bone and Joint Infections: A Systematic Review. Seminars in Nuclear Medicine, 40(1), 3–15. doi:10.1053/j.semnuclmed.2009.08.005

23. Aksoy, S.Y., Asa, S., Ozhan, M. et al. FDG and FDG-labelled leucocyte PET/CT in the imaging of prosthetic joint infection. Eur J Nucl Med Mol Imaging 41, 556–564 (2014). https://doi.org/10.1007/s00259-013-2597-2

24. UNAIDS. Global HIV & AIDS statistics – 2019 fact sheet, July 2019.

25. Sathekge M, Maes A, Kgomo M, et al. Fluorodeoxyglucose uptake by lymph nodes of HIV patients is inversely related to CD4 cell count. Nucl Med Comm 2010;31(2):137e40.

26. Villena García, A. C., Gutierrez Cardo, A., Palomo, L., Hidalgo, C. M., Lillo, E., Espíldora, J., Trigo, J., M., Chaparo, M., A., S., Valdivielso, P. (2019). "18FDG PET/CT & Arterial Inflammation: Predicting Cardiovascular Events in Lung Cancer." QJM: An International Journal of Medicine. doi:10.1093/qjmed/hcz036

27. Keliher, E. J. (2017). Polyglucose nanoparticles with renal elimination and macrophage avidity facilitate PET imaging in ischaemic heart disease. Nature Communications, 8, 14064. doi:10.1038/ncomms14064

28. Sivapackiam, J., Kabra, S., Speidel, S., Sharma, M., Laforest, R., Salter, A., Rettig, M., P., Sharma, V. (2019). 68Ga-Galmydar: A PET imaging tracer for noninvasive detection of Doxorubicin-induced cardiotoxicity. PLOS ONE, 14(5), e0215579. doi:10.1371/journal.pone.0215579

29. Keizer H, Pinedo H, Schuurhuis G, Joenje H. Doxorubicin (adriamycin): a critical review of free radicaldependent mechanisms of cytoxicity. Pharmacol Ther. 1990; 47:219–31. PMID: 2203071

30. Lu, H. D., Wang, L. Z., Wilson, B. K., McManus, S. A., Jumai'an, J., Padakanti, P. K., Alavi, A., Mach, R., H., Prud'homme, R. K. (2018). Copper Loading of Preformed Nanoparticles for PET-Imaging Applications. ACS Applied Materials & Interfaces, 10(4), 3191–3199. doi:10.1021/acsami.7b07242

31. World Health Organization. Global Action Plan on the Public Health Response to Dementia: 2017–2025. Geneva, Switzerland: World Health Organization; 2017.

32. Krasikova, R. (2020). Synthesis and preclinical evaluation of 6-[18F]Fluorine-Alpha-Methyl-L-Tryptophan, a novel PET tracer for measuring tryptophan uptake. ACS Chemical Neuroscience. doi:10.1021/acschemneuro.0c00135

33. Kesler, M., Levine, C., Hershkovitz, D., Mishani, E., Menachem, Y., Lerman, H., Zohar, Y., Shibolet, O., Even-Sapir, E. (2018). 68Ga-PSMA is a novel PET-CT tracer for imaging of hepatocellular carcinoma: A prospective pilot study. Journal of Nuclear Medicine, jnumed.118.214833. doi:10.2967/jnumed.118.214833

34. Mittal S, El-Serag HB. Epidemiology of hepatocellular carcinoma: consider the population. J Clin Gastroenterol. 2013;47(suppl):S2–S6.

35. Hatt, M., Tixier, F., Pierce, L., Kinahan, P. E., Le Rest, C. C., & Visvikis, D. (2016). Characterization of PET/CT images using texture analysis: the past, the present… any future? European Journal of Nuclear Medicine and Molecular Imaging, 44(1), 151–165. doi:10.1007/s00259-016-3427-0

36. Textural Features for Image Classification. Robert M. Haralick, K. Shanmugam, Its'Han Dinstein – 1973.

37. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. Br J Radiol 2017; 90: 20160642.

38. C. Nioche, F. Orlhac, I. Buvat – LIFEx Texture User Guide, LIFEx Version 6.nn – Last update: 2020/08/25.

39. Chicklore, S., Goh, V., Siddique, M., Roy, A., Marsden, P. K., & Cook, G. J. R. (2012). Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. European Journal of Nuclear Medicine and Molecular Imaging, 40(1), 133–140. doi:10.1007/s00259-012-2247-0.

40. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Sci Rep 2016; 6: 34921. doi: https:// doi.org/10.1038/srep34921.

41. van Velden FH, Kramer GM, Frings V, Nissen IA, Mulder ER, de Langen AJ, et al. Repeatability of radiomic features in nonsmall-cell lung cancer [(18)F]FDG-

PET/CT studies: impact of reconstruction and delineation. Mol Imaging Biol 2016; 18: 788–95. doi: https://doi.org/10.1007/ s11307-016-0940-2.

42. Adler S, Seidel J, Choyke P, et al. Minimum lesion detectability as a measure of PET system performance. EJNMMI Phys. 2017;4(1):13. doi:10.1186/s40658-017-0179-2.

43. Leijenaar RT, Nalbantov G, Carvalho S, van Elmpt WJ, Troost EG, Boellaard R, et al. The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep 2015; 5: 11075. doi: https:// doi.org/10.1038/srep11075.

44. Hatt M, Tixier F, Cheze Le Rest C, Pradier O, Visvikis D. Robustness of intratumour ^{18}F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma. Eur J Nucl Med Mol Imaging. 2013;40(11):1662-1671. doi:10.1007/s00259-013-2486-8.

45. Manolekshmi, I. A method for modeling noise in PET images. IOSR Journal of Computer Engineering (IOSR-JCE)e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 6, Ver. II (Nov – Dec. 2014), PP 45-47.

46. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. J Nucl Med. 2020;61(4):488-495. doi:10.2967/jnumed.118.222893.

47. Cook, G.J.R., Siddique, M., Taylor, B.P. et al. Radiomics in PET: principles and applications. Clin Transl Imaging 2, 269–276 (2014). https://doi.org/10.1007/s40336-014-0064-0.

48. Bodalal, Z., Trebeschi, S., Nguyen-Kim, T.D.L. et al. Radiogenomics: bridging imaging and genomics. Abdom Radiol 44, 1960–1984 (2019). https://doi.org/10.1007/s00261-019-02028-w.

49. Nair VS, Gevaert O, Davidzon G, et al. Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. Cancer Res. 2012;72:3725–3734.

50. Yip SSF, Kim J, Coroller TP, et al. Associations between somatic mutations and metabolic imaging phenotypes in non–small cell lung cancer. J Nucl Med. 2017;58:569–576.

51. Arshad MA, Thornton A, Lu H, et al. Discovery of pre-therapy 2-deoxy-2-18Ffluoro-D-glucose positron emission tomography-based radiomics classifiers

of survival outcome in non-small-cell lung cancer patients. Eur J Nucl Med Mol Imaging. 2019;46:455–466.

52. Jiang Y, Yuan Q, Lv W, et al. Radiomic signature of 18F fluorodeoxyglucose PET/CT for prediction of gastric cancer survival and chemotherapeutic benefits. Theranostics. 2018;8:5915–5928.

53. Radon, J. On the determination of functions from their integral values along certain manifolds. IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. MI-5, NO. 4, DECEMBER 1986.

54. Jones, Debra & Cookson, James. (2010). Johnson Matthey Academic Conference 2009. Platinum Metals Review - PLATIN MET REV. 54. 56-60. 10.1595/147106709X481831.

55. DHD Multimedia Gallery.

56. Shahhosseini S. PET Radiopharmaceuticals. Iran J Pharm Res. 2011 Winter;10(1):1-2. PMID: 24363674; PMCID: PMC3869596.

57. Yao R, Lecomte R, Crawford ES. Small-animal PET: what is it, and why do we need it? [published correction appears in J Nucl Med Technol. 2013 Jun;41(2):118. Dosage error in article text]. J Nucl Med Technol. 2012;40(3):157-165. doi:10.2967/jnmt.111.098632.

58. Alsyed, Emad & Smith, Rhodri & Marshall, Christopher & Spezi, Emiliano. (2019). EP-0811 Stability of PET Radiomic Features: A Preclinical Study.

59. Orlhac F, Thézé B, Soussan M, Boisgard R, Buvat I. Multiscale Texture Analysis: From 18F-FDG PET Images to Histologic Images. J Nucl Med. 2016;57(11):1823-1828. doi:10.2967/jnumed.116.173708.

60. Holbrook MD, Blocker SJ, Mowery YM, Badea A, Qi Y, Xu ES, Kirsch DG, Johnson GA, Badea CT. MRI-Based Deep Learning Segmentation and Radiomics of Sarcoma in Mice. Tomography. 2020 Mar;6(1):23-33. doi: 10.18383/j.tom.2019.00021. PMID: 32280747; PMCID: PMC7138523.

61. Veres, D.S., Máthé, D., Hegedűs, N. et al. Radiomic detection of microscopic tumorous lesions in small animal liver SPECT imaging. EJNMMI Res 9, 67 (2019). https://doi.org/10.1186/s13550-019-0532-7.

62. John Hopkins Medicine - https://www.hopkinsmedicine.org/

63. Catalina Imaging - https://catalinaimaging.com/

64. Boschi, A.; Uccelli, L.; Martini, P. A Picture of Modern Tc-99m Radiopharmaceuticals: Production, Chemistry, and Applications in Molecular Imaging. Appl. Sci. 2019, 9, 2526.

65. Podoloff, Donald & Advani, Ranjana & Allred, Craig & Benson, Al & Brown, Elizabeth & Burstein, Harold & Carlson, Robert & Coleman, R & Czuczman, Myron & Delbeke, Dominique & Edge, Stephen & Ettinger, David & Grannis, Frederic & Hillner, Bruce & Hoffman, John & Kiel, Krystyna & Komaki, Ritsuko & Larson, Steven & Mankoff, David & Zelenetz, Andrew. (2007). NCCN Task Force report: Positron Emission Tomography (PET)/Computed Tomography (CT) scanning in cancer. Journal of the National Comprehensive Cancer Network : JNCCN. 5 Suppl 1. S1-22; quiz S23. 10.6004/jnccn.2007.2001.

66. Lee, Y.-B., Choi, K.M. Diet-Modulated Lipoprotein Metabolism and Vascular Inflammation Evaluated by 18F-fluorodeoxyglucose Positron Emission Tomography. Nutrients 2018, 10, 1382.

67. Chryssikos, Timothy & Parvizi, Javad & Ghanem, Elie & Newberg, Andrew & Zhuang, Hongming & Alavi, Abass. (2008). FDG-PET Imaging Can Diagnose Periprosthetic Infection of the Hip. Clinical orthopaedics and related research. 466. 1338-42. 10.1007/s11999-008-0237-0.

68. Längkvist, Martin & Alirezaie, Marjan & Kiselev, Andrey & Loutfi, Amy. (2016). Interactive Learning with Convolutional Neural Networks for Image Labeling.

69. Hinzpeter, Ricarda & Wagner, Matthias & Wurnig, Moritz & Seifert, Burkhardt & Manka, Robert & Alkadhi, Hatem. (2017). Texture analysis of acute myocardial infarction with CT: First experience study. PLOS ONE. 12. e0186876. 10.1371/journal.pone.0186876.

70. Scrivener, Madeleine & de Jong, Evelyn & Van Timmeren, Janita & Pieters, Thierry & Ghaye, Benoît & Geets, Xavier. (2016). Radiomics applied to lung cancer: A review. Translational Cancer Research. 5. 398-409. 10.21037/tcr.2016.06.18.

71. Jansen, Robin & Amstel, Paul & Martens, Roland & Kooi, Irsan & Wesseling, Pieter & de Langen, Adrianus Johannes & Menke-van der Houven van Oordt, Catharina & Jansen, Bernard & Moll, Annette & Dorsman, Josephine & Castelijns, Jonas & Graaf, Pim & de Jong, Marcus. (2018). Non-invasive tumor

genotyping using radiogenomic biomarkers, a systematic review and oncology-wide pathway analysis. Oncotarget. 9. 10.18632/oncotarget.24893.

72. Youn, Hyewon & Hong, Kee-Jong. (2012). In vivo Noninvasive Small Animal Molecular Imaging. Osong public health and research perspectives. 3. 48-59. 10.1016/j.phrp.2012.02.002.

73. M. Magdics, B. Tóth, B. Kovács, et al. (2011) Total variation regularization in PET reconstruction, Képfeldolgozók és Alakfelismerók VIII. Konferenciája. Szeged.

74. Magdics M, Szirmay-Kalos L, Szlavecz A, et al. (2010) TeraTomo project: a fully 3D GPU based reconstruction code for exploiting the imaging capability of the NanoPET™/CT system. Mol Imaging Biol 12.

75. Nagy K, Toth M, Major P et al. (2013) Performance evaluation of the small-animal nanoScan PET/MRI system. J Nucl Med 54:18125– 11832.

76. A. Gaitanis, G. A. Katsis et al. (2017) Investigation of Image Reconstruction Parameters of the Mediso nanoScan PC Small-Animal PET/CT Scanner for Two Different Positron Emitters Under NEMA NU 4-2008 Standards. Mol Imaging Biol (2017) 19:550Y559. DOI: 10.1007/s11307-016-1035-9.

77. Hatzioannou A, Nayar S, Gaitanis A, Barone F, Anagnostopoulos C, Verginis P. Intratumoral accumulation of podoplanin-expressing lymph node stromal cells promote tumor growth through elimination of CD4+ tumor-infiltrating lymphocytes. Oncoimmunology. 2016;5(9):e1216289. Published 2016 Jul 29. doi:10.1080/2162402X.2016.1216289.

78. National Electrical Manufacturers Association (NEMA) (2008) Performance measurements for small animal positron emission tomographs (PETs). NEMA Standards Publication NU 4-2008. Rosslyn, VA: NEMA.

79. C Nioche, F Orlhac, S Boughdad, S Reuzé, J Goya-Outi, C Robert, C Pellot-Barakat, M Soussan, F Frouin, and I Buvat. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Research 2018; 78(16):4786-4789 (www.lifexsoft.org).

80. Branchini M, Zorz A, Zucchetta P, et al. Impact of acquisition count statistics reduction and SUV discretization on PET radiomic features in pediatric 18F-FDG-

PET/MRI examinations. Phys Med. 2019;59:117-126. doi:10.1016/j.ejmp.2019.03.005.

81. Nyflot MJ, Yang F, Byrd D, Bowen SR, Sandison GA, Kinahan PE. Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards. J Med Imaging (Bellingham). 2015;2(4):041002. doi:10.1117/1.JMI.2.4.041002.

82. Doyle, Laurance & McCowan, Brenda & Hanser, Sean & Christopher, Chyba & Taylor, Bucci & Blue, J. (2008). Applicability of Information Theory to the Quantification of Responses to Anthropogenic Noise by Southeast Alaskan Humpback Whales. Entropy: International and Interdisciplinary Journal of Entropy and Information Studies. 10. 10.3390/entropy-e10020033.

83. OsiriX: An Open-Source Software for Navigating in Multidimensional DICOM Images.J Digit Imaging. 2004 Sep;17(3):205-216. – https://www.osirix-viewer.com/

84. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine. 2016; 15(2):155–63. https://doi.org/10.1016/j.jcm.2016.02. 012 PMID: 27330520; PubMed Central PMCID: PMC4913118.

85. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. Acta Oncol. 2010;49(7):1012-1016. doi:10.3109/0284186X.2010.498437.

86. Foley, K.G., Hills, R.K., Berthon, B. et al. Development and validation of a prognostic model incorporating texture analysis derived from standardised segmentation of PET in patients with oesophageal cancer. Eur Radiol 28, 428–436 (2018). https://doi.org/10.1007/s00330-017-4973-y.

87. Feliciani G, Fioroni F, Grassi E, et al. Radiomic Profiling of Head and Neck Cancer: 18F-FDG PET Texture Analysis as Predictor of Patient Survival. Contrast Media & Molecular Imaging. 2018 ;2018:3574310. DOI: 10.1155/2018/3574310.

88. Nakajo, M., Jinguji, M., Aoki, M. et al. The clinical value of texture analysis of dual-time-point 18F-FDG-PET/CT imaging to differentiate between 18F-FDG-avid

benign and malignant pulmonary lesions. Eur Radiol 30, 1759–1769 (2020). https://doi.org/10.1007/s00330-019-06463-7.

89. Ralph T. H. Leijenaar, Sara Carvalho, Emmanuel Rios Velazquez, Wouter J. C. van Elmpt, Chintan Parmar, Otto S. Hoekstra, Corneline J. Hoekstra, Ronald Boellaard, André L. A. J. Dekker, Robert J. Gillies, Hugo J. W. L. Aerts & Philippe Lambin (2013) Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability, Acta Oncologica, 52:7, 1391-1397, DOI: 10.3109/0284186X.2013.812798.

90. Van Velden, F. H. P., Nissen, I. A., Jongsma, F., Velasquez, L. M., Hayes, W., Lammertsma, A. A., Hoekstra, O., S., Boellaard, R. (2013). Test-Retest Variability of Various Quantitative Measures to Characterize Tracer Uptake and/or Tracer Uptake Heterogeneity in Metastasized Liver for Patients with Colorectal Carcinoma. Molecular Imaging and Biology, 16(1), 13–18. doi:10.1007/s11307-013-0660-9.

91. Cheng, N., Hsieh, C., Fang, Y.D. et al. Development and validation of a prognostic model incorporating [18F]FDG PET/CT radiomics for patients with minor salivary gland carcinoma. EJNMMI Res 10, 74 (2020). https://doi.org/10.1186/s13550-020-00631-3.

92. Ger RB, Meier JG, Pahlka RB, Gay S, Mumme R, Fuller CD, et al. (2019) Effects of alterations in positron emission tomography imaging parameters on radiomics features. PLoS ONE 14(9): e0221877. https://doi.org/10.1371/journal.pone.0221877.

93. Chapter 2.n Summarizing Data: listing and grouping - Statistics. Provided by: Wikidot. Located at: http://statistics.wikidot.com/ch2.