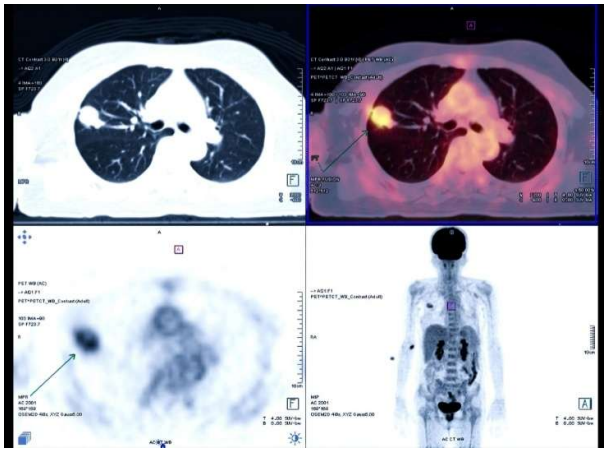
POSITRON EMISSION TOPOGRAPHY (PET)

-Aarushi Sharma , Educational Content Writer, MacroEdtech

Introduction to PET

Positron Emission Tomography (PET) is a type of nuclear medicine procedure that measures metabolic activity of the cells of body tissues. It is a combination of nuclear medicine and biochemical analysis tools.



The process involves an injection of a safe radioactive tracer that helps detect diseased cells. These cells absorb large amounts of radiotracers, which indicates potential health problems by showing both atypical and typical metabolic activity. Healthcare providers frequently use PET scans to help diagnose cancer and assess cancer treatment. They can also assess certain heart and brain issues with the scan. For example, in PET scans of the brain, a radioactive atom is applied to glucose (blood sugar) to create a radionuclide called Fluorodeoxyglucose(FDG).

Comparatively, Computed Tomography (CT) scans uses X-rays, while Magnetic Resonance Imaging (MRI) scans use magnets and radio waves. Both produce still images of organs and body structures. But PET scans use a radioactive tracer to show how an organ is functioning in real time. PET scan images can detect cellular changes in organs and tissues earlier than CT and MRI scans, by detecting the atypical metabolism before the disease shows up on the imaging tests.

It differs from other nuclear medicine examinations in that PET detects metabolism within body tissues, whereas other types of nuclear medicine examinations detect the amount of a radioactive substance collected in body tissue in a certain location to examine the tissue's function.

The defining strength of PET lies in its ability to perform functional and molecular imaging. PET tracers are designed to mimic naturally occurring biological molecules such as glucose, amino acids or neurotransmitters. When introduced into the body, these tracers participate in physiological pathways, allowing PET to visualise cellular processes rather than static structures.

This capability enables PET to detect disease at a stage when molecular dysfunction has begun but structural changes are minimal or absent. As a result, PET is uniquely suited for early disease detection, assessment of disease activity and evaluation of treatment response at a biochemical level.

Today, PET plays a central role in oncology, cardiology and neurology, as well as in infection imaging and research applications. In oncology, PET is indispensable for tumour staging, treatment planning and response assessment. In cardiology, it allows evaluation of myocardial viability and perfusion. In neurology, PET provides insights into epilepsy, neurodegenerative disorders and brain tumours.

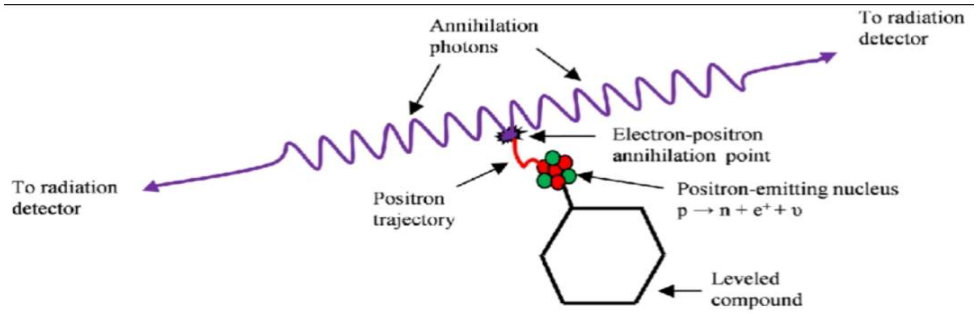
Beyond diagnosis, PET has become a cornerstone of precision medicine, guiding targeted therapies and enabling personalised treatment strategies. As tracer development, detector technology and computational methods continue to evolve, PET is increasingly positioned not merely as an imaging tool, but as a window into the dynamic molecular processes that define health and disease.

Physics Of PET Imaging

PET Imaging begins with the injection of a metabolically active tracer biological molecule that carries with it a positron-emitting isotope (for example, ^{11}C, ^{13}N, ^{15}O or ^{18}F).

Over a few minutes, the isotope accumulates in an area of the body for which the molecule has an affinity. For example, glucose labelled with ^{11}C, or a glucose analog labelled with ^{18}F, accumulates in the brain or in tumours, where glucose is used as the primary source of energy. The radioactive nuclei then decay by positron emission. The ejected positron combines with an electron almost instantaneously, and these two particles undergo the process of annihilation. The energy associated with the masses of the positron and electron divided equally between two photons that fly away from one another at a 180° angle. Each photon has an energy of 511 keV. These high-energy gamma rays emerge from the body in opposite directions, to be detected by an array of detectors that surround the patient.

When two photons are recorded simultaneously by a pair of detectors, the annihilation event that gave rise to them must have occurred somewhere along the line connecting the detectors. Of course, if one of the photons is scattered, then the line of coincidence will be incorrect. After 100,000 or more annihilation events are detected, the distribution of the positron-emitting tracer is calculated by the tomographic reconstruction procedures from the recorded projection data.



Time-of-flight is an advanced technique in modern medicine, which improves image quality and scanner efficiency by measuring the precise time difference (Δt) between the detection of two annihilation photons, narrowing down the emission location along the line-of-response (LOR). This reduces image noise and increases signal-to-noise ratio (SNR), providing faster scans and better contrast.

Principle: When a positron annihilates, it emits two 511-keV photons in opposite directions. While conventional PET only identifies the LOR, TOF-PET measures the tiny time difference (t1-t2) to pinpoint the specific location of annihilation, calculated as

$$\Delta x = (c \times t)/2$$

Where c is speed of light.

Advantages of this technique include:

- **Improved Accuracy:** the technique localises the event within a smaller region along the LOR, which reduces the number of pixels considered during image reconstruction, this improving the SNR.
- **Technology:** TOF-PET requires ultra-fast scintillators and detectors capable of precise timing revolution (eg. sub-nanosecond range) to distinguish which photon arrived first.

- **Benefits**: the primary advantages include enhanced image contrast, reduced artifacts, and better lesion detection, which are particularly beneficial for identifying small or low-contrast tumors.

Hence, simply put, time-of-flight offers higher image quality for more accurate detection of masses and lesions, especially in large patients. A more accurate system gives you better information, and better information makes it easier to declare a more definitive diagnosis and pursue a more specialized treatment plan.

The superior sensitivity of Positron Emission Tomography (PET) compared to Single Photon Emission Computed Tomography (SPECT) arises from fundamental differences in how emitted radiation is detected, rather than from better tracers or stronger scanners alone.

Fundamental Difference in Detection Geometry

In SPECT, gamma photons emitted from the radiotracer travel in all directions. To determine the direction of origin, physical collimators made of lead are placed in front of the detector. These collimators allow only photons traveling in specific directions to reach the detector, while the vast majority are absorbed.

As a result, more than 99% of emitted photons are discarded before detection.

In contrast, PET does not require physical collimation. Instead, PET relies on electronic collimation. When a positron emitted from the radionuclide annihilates with an electron, it produces two 511 keV photons emitted simultaneously in nearly opposite directions (≈180°). PET scanners detect these photons in coincidence, defining a line of response (LOR) between the detectors.

Because PET determines photon direction electronically rather than mechanically, a much larger fraction of emitted radiation contributes to image formation, resulting in substantially higher sensitivity.

Coincidence Detection Eliminates the Need for Collimators

Collimators are the primary cause of sensitivity loss in SPECT. PET eliminates this trade-off. By requiring simultaneous detection of two photons, PET intrinsically localizes the annihilation event along a line without absorbing photons unnecessarily.

PET Scanner Design & Image Formation

Detector Crystals

At the heart of a PET scanner are scintillation detector crystals, which convert incoming gamma photons into visible light. Modern PET systems predominantly use high-density, high-atomic-number crystals such as Bismuth Germanate (BGO), Lutetium Oxyorthosilicate (LSO) and Lutetium Yttrium Oxyorthosilicate (LYSO).

These materials are selected because 511 keV annihilation photons require detectors with strong stopping power. LSO and LYSO, in particular, offer a combination of high density, fast scintillation decay times and high light output, making them ideal for modern high-count-rate imaging. Faster decay times allow detectors to distinguish closely spaced events, improving timing resolution and enabling advanced techniques such as time-of-flight (TOF) PET. BGO, while having lower light output, is still used in some systems due to its excellent stopping efficiency.

The choice of crystal directly affects sensitivity, timing accuracy and overall image noise.

Photodetectors: PMTs vs SiPMs

Once scintillation light is produced, it must be converted into an electrical signal. Traditionally, PET scanners used photomultiplier tubes (PMTs), which amplify light signals through a cascade of electron multiplication. PMTs are highly sensitive but bulky, fragile and sensitive to magnetic fields.

Modern PET systems increasingly use silicon photomultipliers (SiPMs). These solid-state devices offer several advantages:

- Higher timing resolution
- Compact size
- Resistance to magnetic fields
- Compatibility with PET/MRI systems

SiPMs allow for more compact detector designs, better timing precision and improved signal detection, which collectively enhance image quality and quantitative accuracy.

Ring Geometry and Data Acquisition

PET scanners are designed as circular or cylindrical rings of detectors surrounding the patient. When a positron annihilation event occurs, the two emitted photons are detected simultaneously by opposing detectors. This coincidence detection defines a line of response (LOR), along which the annihilation event must have occurred.

Modern PET scanners collect millions of such LORs during an acquisition. Events are classified as:

- True coincidences (both photons from the same annihilation)
- Scatter coincidences (photons deflected before detection)
- Random coincidences (unrelated photons detected simultaneously)

Sophisticated electronics and timing windows are used to maximize true events while minimizing noise. In time-of-flight PET, the slight difference in arrival times of the two photons is used to further localize the annihilation point along the LOR, improving image contrast and reducing noise.

Image Reconstruction

The raw coincidence data collected by the scanner does not directly form an image. Instead, it must undergo image reconstruction, which mathematically converts detected events into a three-dimensional representation of tracer distribution.

Early PET systems used analytical methods such as filtered back projection. Modern PET scanners rely on iterative reconstruction algorithms, such as Ordered Subsets Expectation Maximization (OSEM). These methods repeatedly refine the image by comparing measured data with estimated tracer distributions, incorporating corrections for attenuation, scatter, random events and detector response.

Iterative reconstruction improves image quality, contrast and quantification, especially in low-count studies, making PET images more reliable for clinical interpretation.

Radiopharmaceuticals in PET

Suitable PET radionuclide is defined by its ability to produce a high-resolution, quantitative images of biological processes while minimizing radiation dose to the patient. The most critical characteristics determining suitability are a favourable decay profile, an appropriate half-life, and the ability to be incorporated into a biologically active molecule.

Key Characteristics

- Decay Profile (Low Energy Positrons): An ideal radionuclide should emit positrons with low kinetic energy. Lower energy means a shorter travel distance(linear range) in tissue before annihilation, resulting in higher spatial resolution.

- Half-life Matching Biology: The radionuclide's half-life should match the biological process or pharmacokinetics of the tracer, allowing enough time for accumulation in target tissues and clearance from non-target areas.
- High Branching Ratio: A high positron emission branching ratio (β^+ decay) is essential to maximize the signal and reduce the required radiation dose.
- Chemical Versatility: The isotope should be easily incorporated into compounds, such as glucose analogues (eg. FDG), water or receptor ligand.
- Accessibility and Production: It should be producible by a cyclotron or, in some cases, a generator in high purity.

Depending on the targeted processes within the living organisms , different tracers are used for various medical conditions, such as cancer, particular brain pathologies, cardiac events and bone lesions, where the most commonly used tracers are radiolabelled with ^{18}F (eg. [^{18}F]- FDG and NA[^{18}F]).

A radiopharmaceutical compound consists of: 1) a molecular structure identified as a vehicle molecule and 2) a positron-emitting radionuclide. The radioisotope is attached to the vehicle molecule, also known as a ligand, and then injected into the body as a radioactive tracer.

Oxygen-15 isotope is mostly involved in blood flow measurements , whereas a wide array of ^{11}C-based compounds have also been developed for neuronal disorders according to the affected neuroreceptors, prostate cancer, and lung carcinomas.

In contrast, the single-photon emission computed tomography (SPECT) technique uses gamma-emitting radioisotopes and can be used to diagnose strokes, seizures, bone illnesses, and infections by gauging the blood flow and radio within tissues and organs. The radioisotopes typically used in SPECT imaging are Iodine-123, Technetium-99, Xenon-133, Thallium-201 and Indium-111.

Commonly used PET isotopes include Fluorine-18, Carbon-11, Nitrogen-13, Oxygen-15 and Gallium-68. These radionuclides have short half-lives, necessitating on-site cyclotrons for ^{11}C, ^{13}N and ^{15}O, while ^{18}F and ^{68}Ga allow for transport or generator production.

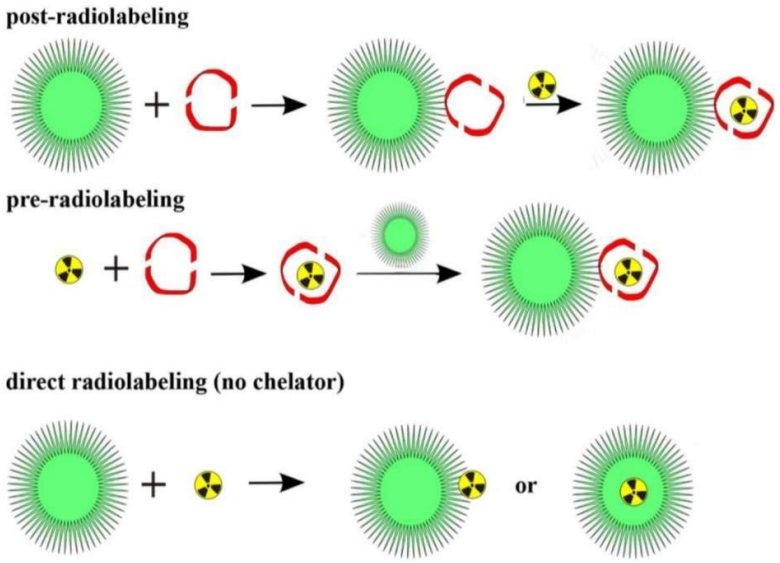
Radiolabelling is the process by which a radioactive isotope is chemically attached to a biologically active molecule, forming a radiopharmaceutical capable of tracing physiological processes in vivo. The success of PET imaging depends not only on the physical properties of the radionuclide but also on the stability, specificity and purity of this labelled compound. Effective radiolabelling ensures that the detected signal

accurately represents biological function rather than artefacts arising from tracer breakdown or nonspecific distribution.

Principles of Radiolabelling

A fundamental requirement of radiolabelling is chemical stability. The radioactive isotope must remain firmly bound to its carrier molecule throughout its circulation in the body. If the radionuclide dissociates, it may accumulate in non-target tissues, leading to misleading images and increased radiation dose to unintended organs. To prevent this, isotopes are attached using chelators, covalent bonds or coordination complexes designed to withstand physiological conditions.

Another critical principle is half-life matching. The physical half-life of the radionuclide must be compatible with the biological half-life of the tracer. Short-lived isotopes such as F-18 are ideal for metabolic tracers because they provide sufficient imaging time while minimizing radiation exposure. Longer-lived isotopes are reserved for slower biological processes. A mismatch between physical decay and biological behaviour can result in poor image quality or unnecessary patient dose.



Radiolabelling must also preserve the biological activity of the molecule. The chemical modification introduced during labelling should not alter receptor binding, enzyme interaction or transport mechanisms. Even small structural changes can significantly affect bio-distribution, making it essential that the labelled compound behaves identically to its non-radioactive counterpart.

Equally important is specific activity, defined as the radioactivity per unit mass of the compound. High specific activity ensures that tracer doses remain low enough to avoid pharmacological effects while still providing sufficient signal for detection. This is

particularly crucial in receptor-based imaging, where excessive non-radioactive ligand could saturate target sites and distort results.

Quality Control and Purity

Once radiolabelling is complete, rigorous quality control is essential before clinical use. Unlike conventional pharmaceuticals, radiopharmaceuticals cannot rely on long-term stability testing due to radioactive decay, making rapid and reliable quality checks critical.

Radiochemical purity is one of the most important parameters. It refers to the proportion of radioactivity present in the desired chemical form. Impurities may include free radionuclide, degraded products or incorrectly labelled compounds. Techniques such as thin-layer chromatography (TLC) and high-performance liquid chromatography (HPLC) are routinely used to confirm that the radiotracer meets acceptable purity standards.

Chemical purity ensures that residual solvents, unreacted precursors or toxic reagents used during synthesis are below safe limits. Even trace contaminants can affect patient safety, particularly because radiopharmaceuticals are administered intravenously.

Sterility and pyrogen testing are mandatory, as radiopharmaceuticals bypass normal physiological barriers. Although full sterility testing may be performed retrospectively due to time constraints, aseptic preparation techniques and validated synthesis protocols are essential to minimize risk.

pH and isotonicity are also evaluated to ensure patient comfort and compatibility with blood chemistry. Deviations can cause pain, tissue irritation or hemolysis.

Biological Basis of PET Imaging

The true power of PET lies in its ability to exploit normal cellular biology for diagnostic purposes. PET radiotracers are designed to participate in physiological pathways rather than merely accumulate passively. Once administered, a tracer's distribution is governed by tracer kinetics: delivery via blood flow, transport across cell membranes, intracellular metabolism or binding, and eventual clearance. This bio-distribution is not random. It reflects the expression of transporters, the activity of enzymes, and the density of receptors within tissues. As a result, PET does not simply show where a tracer is located, but why it is there, providing a quantitative map of biological function rather than anatomy alone.

Cancer cells unintentionally expose themselves to PET imaging because of their altered metabolism. Most malignant cells exhibit the Warburg effect, a preference for aerobic glycolysis even in the presence of adequate oxygen. This metabolic reprogramming leads

to markedly increased glucose uptake and utilization. Fluorodeoxyglucose (FDG), a glucose analogue, exploits this vulnerability. FDG enters cells via overexpressed glucose transporters and is phosphorylated by hexokinase, but cannot proceed further through glycolysis. As a result, it becomes metabolically trapped within the cell. Regions of high FDG accumulation therefore represent areas of increased metabolic demand, allowing tumours to appear as intense focal signals on PET images, often before any structural abnormality becomes apparent.

Beyond glucose metabolism, PET tracers can be engineered to target specific receptors, enzymes, or cellular processes, extending PET imaging far beyond oncology. Hormone receptors, amino acid transporters, hypoxia markers, and neurotransmitter systems can all be selectively visualised. Importantly, PET distinguishes physiological uptake from pathological uptake through characteristic distribution patterns and quantitative analysis. Normal tissues such as the brain, myocardium, and kidneys demonstrate predictable tracer accumulation based on their intrinsic metabolic roles. Pathological uptake, by contrast, reflects dysregulated biology, uncontrolled proliferation, aberrant receptor expression, or altered enzymatic activity. In this way, cancer cells are not actively "detected" by PET; rather, they reveal themselves by amplifying the very biological processes required for their survival.

Quantification in PET

While PET images are often visually interpreted, their true clinical and scientific value lies in quantification. Unlike conventional imaging, PET allows tracer uptake to be measured numerically, enabling objective assessment of disease activity, comparison across time points, and evaluation of treatment response. Quantitative PET transforms images from qualitative "hot spots" into measurable biological data.

Standard Uptake Value (SUV)

The most widely used quantitative metric in PET is the Standard Uptake Value (SUV). SUV represents the concentration of radiotracer within a region of interest, normalized to the injected dose and the patient's body parameters, most commonly body weight. In practice, SUV provides a semi-quantitative estimate of tracer uptake, allowing lesions to be compared within the same patient or across serial scans. Variants such as SUVmax, SUVmean, and SUVpeak are used depending on clinical context, with SUVmax being particularly useful due to its reproducibility and relative independence from region-of-interest definition.

Factors Affecting SUV Accuracy

Despite its widespread use, SUV is influenced by numerous technical and biological factors. Patient-related variables such as body composition, blood glucose levels, and time elapsed between tracer injection and imaging can significantly alter measured uptake. Technical factors including scanner calibration, reconstruction algorithms, attenuation correction, and region-of-interest selection further contribute to variability. As a result, SUVs must be interpreted within a standardized imaging protocol, and comparisons are most reliable when acquisition and processing parameters are kept consistent across studies.

Dynamic PET and Kinetic Modelling

Beyond static SUV measurements, dynamic PET imaging offers a more rigorous approach to quantification. Dynamic PET involves continuous or sequential imaging immediately following tracer injection, generating time–activity curves for tissues of interest. These curves can be analysed using kinetic models to estimate physiological parameters such as transport rates, binding potentials, or metabolic flux. While more complex and time-intensive, kinetic modelling provides a more accurate representation of tracer behaviour and biological function, particularly in research settings and advanced clinical applications.

Partial Volume Effect

Quantification in PET is also limited by the partial volume effect, which arises from the finite spatial resolution of PET scanners. When structures are smaller than approximately two to three times the system's spatial resolution, measured activity is underestimated due to signal spill-out into surrounding tissue. This effect is especially relevant for small lesions, thin cortical structures, and paediatric imaging. Partial volume correction techniques exist, but they require accurate anatomical information and introduce additional complexity.

Limitations of PET Quantification

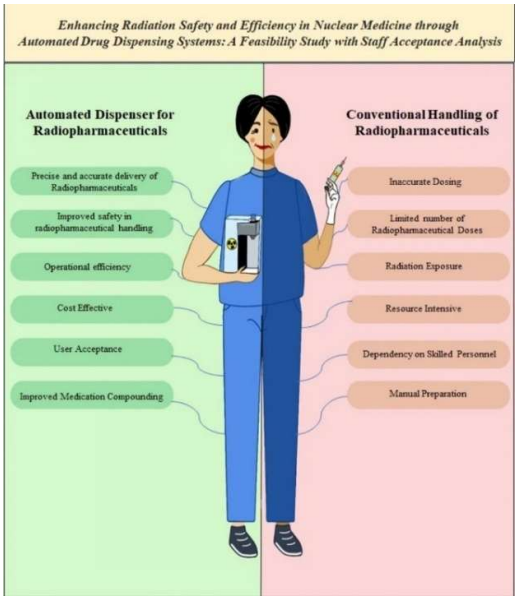
Although PET offers powerful quantitative capabilities, it is not a purely absolute measurement technique. Variability in biological conditions, scanner performance, and image processing limits the precision of quantification. Additionally, many PET tracers reflect multiple biological processes simultaneously, complicating interpretation. Consequently, PET quantification is most effective when used in conjunction with clinical context, anatomical imaging, and longitudinal comparison rather than as an isolated numerical value.

Safety, Dosimetry and Limitations

While PET offers unparalleled insight into molecular and functional processes, its use is inseparable from considerations of radiation safety, dosimetry, and practical limitations. Responsible application of PET requires balancing diagnostic benefit against radiation exposure, particularly in vulnerable populations and in repeated imaging scenarios.

For patients, PET exposure is episodic and clinically justified, with the diagnostic or therapeutic benefit far outweighing the associated risk. In contrast, occupational exposure concerns healthcare workers involved in radiopharmaceutical production, handling, and administration. Strict radiation protection measures, such as shielding, time minimization, distance optimization, and automated dispensing systems, ensure that occupational doses remain well below regulatory limits. Modern nuclear medicine practice has significantly reduced staff exposure through improved workflow design and technology.

Special caution is required when considering PET imaging in pregnant patients due to potential foetal radiation exposure, particularly during early gestation. PET is generally avoided in pregnancy unless the clinical benefit is critical and no suitable alternatives exist. In paediatric imaging, radiation sensitivity is higher due to rapidly dividing tissues



and longer life expectancy. Paediatric PET protocols therefore employ lower administered activities, weight-based dosing, and stringent justification to minimize risk while preserving diagnostic value.

Despite its strengths, PET has inherent limitations. High operational costs related to cyclotron production, radiochemistry infrastructure, and scanner technology restrict availability, particularly in resource-limited settings. PET also suffers from limited spatial resolution compared to CT or MRI, making precise anatomical localization challenging without hybrid imaging. False-positive findings can occur due to physiological or inflammatory tracer uptake, necessitating careful interpretation within clinical context. Additionally, the short half-lives of many PET isotopes impose logistical constraints on production, transport, and scheduling.

Advances and Future of PET

Positron Emission Tomography is undergoing rapid transformation driven by advances in detector technology, radiochemistry and computational methods. These developments are expanding PET beyond conventional imaging toward a role as a central tool in precision and personalized medicine. The future of PET is defined not only by improved image quality, but by deeper biological insight and more individualized clinical decision-making.

One of the most significant recent advances is the development of total-body PET scanners, which extend detector coverage to nearly the entire length of the human body. Unlike conventional PET systems that image a limited axial field of view, total-body PET can capture tracer distribution across all organs simultaneously. This dramatically increases sensitivity, allowing for ultra-low dose imaging, faster acquisition times, and the ability to study whole-body tracer kinetics in real time. Total-body PET also enables new applications, such as tracking systemic disease processes, immune responses, and inter-organ metabolic interactions that were previously inaccessible.

Artificial intelligence is increasingly integrated into PET image reconstruction and analysis. AI-based reconstruction algorithms can reduce noise, correct motion artefacts, and enhance image resolution without increasing radiation dose. Machine learning models are also being developed to standardize quantitative measurements, reduce inter-scanner variability, and assist in automated lesion detection. By improving consistency and efficiency, AI has the potential to make PET imaging more accessible and reproducible across clinical settings.

A major paradigm shift in nuclear medicine is the rise of **theranostics**, where diagnostic imaging and targeted therapy are linked through the same molecular pathway. PET plays a central role in this approach by identifying patients whose tumours express specific targets and predicting therapeutic response. Diagnostic PET tracers guide the selection of corresponding therapeutic radionuclides, enabling individualized treatment strategies. This integration of imaging and therapy exemplifies the movement toward precision oncology.

Together, these advances position PET as a cornerstone of personalized imaging. Rather than providing a one-size-fits-all assessment, PET can be tailored to the biological characteristics of each patient's disease. By quantifying molecular activity, predicting treatment response, and monitoring therapy at an early stage, PET supports clinical decisions that are increasingly patient-specific rather than protocol-driven.

Data Analysis for Educational purpose

By Aarushi Sharma

1. Detailed Description of PET _Cancer Detection Data

Positron Emission Tomography (PET) imaging is a functional diagnostic technique widely used in oncology to detect cancerous tissues based on their metabolic activity. In PET scans, a radioactive tracer such as Fluorodeoxyglucose (FDG) is injected into the patient's body. Since cancer cells typically exhibit higher metabolic rates than normal cells, they absorb more FDG, which results in higher signal intensity in PET images. The PET machine measures the distribution and concentration of this tracer across different body regions, producing quantitative metrics such as Standardized Uptake Value (SUV), lesion size, metabolic tumor volume, and total lesion glycolysis. The dataset generated from PET machines consists of patient-level and lesion-level observations, including physiological parameters (age, weight, blood glucose), imaging parameters (SUVmax, SUVmean, scan duration), anatomical information (organ or body region), and diagnostic labels indicating whether the detected region is malignant, benign, or normal. Such data is crucial for early cancer detection, staging, treatment planning, therapy response monitoring, and prognosis evaluation. Large datasets with thousands of readings help in building machine learning models that assist radiologists in identifying suspicious lesions, reducing diagnostic errors, and improving clinical decision-making.

2. Objective of Data Analysis

The main objectives of analyzing PET cancer data are:

Identify patterns of high metabolic activity associated with malignant tumors

Differentiate malignant vs benign vs normal lesions

Detect threshold SUV values that indicate cancer risk

Study relationships between patient factors (age, glucose level, weight) and cancer detection

Build predictive models for early cancer detection

Evaluate tumor burden using metabolic volume and lesion size

Support clinical decision-making and automated diagnostic systems

Key interpretations you should derive:

Higher SUVmax → higher probability of malignancy

Larger lesion size + high SUV → aggressive tumor indication

High blood glucose may affect PET accuracy

Organ-specific cancer distribution patterns

Probability-based risk scoring for early detection

```
import pandas as pd
import numpy as np
import random

from google.colab import files
uploaded = files.upload()

#View Data
df.head(20)

df.describe()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving PET_Cancer_Data_12000.csv to PET_Cancer_Data_12000 (5).csv
```

	Patient_ID	Age	Weight_kg	Blood_Glucose_mg_dL	Injected_Dos
count	12000.00000	12000.000000	12000.000000	12000.000000	12000.0
mean	6000.50000	52.078083	70.124283	100.003292	370.2
std	3464.24595	18.722291	11.995175	19.764212	50.1
min	1.00000	20.000000	26.600000	16.300000	183.4
25%	3000.75000	36.000000	61.900000	86.300000	335.7
50%	6000.50000	52.000000	70.100000	99.900000	370.0
75%	9000.25000	68.000000	78.300000	113.400000	404.4

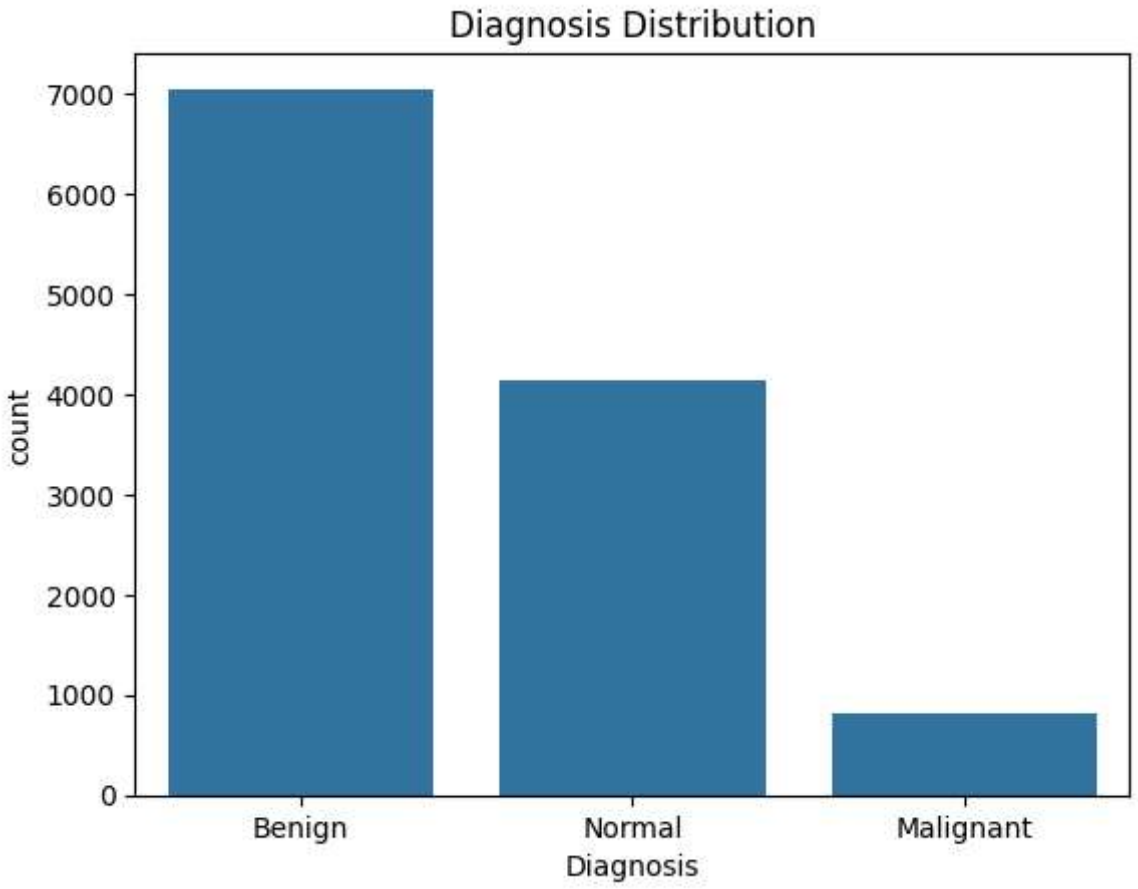
```
#Check Missing Values
df.isnull().sum()

#Diagnosis Distribution
df["Diagnosis"].value_counts()
```

Diagnosis	count
Benign	7049
Normal	4143
Malignant	808

dtype: int64

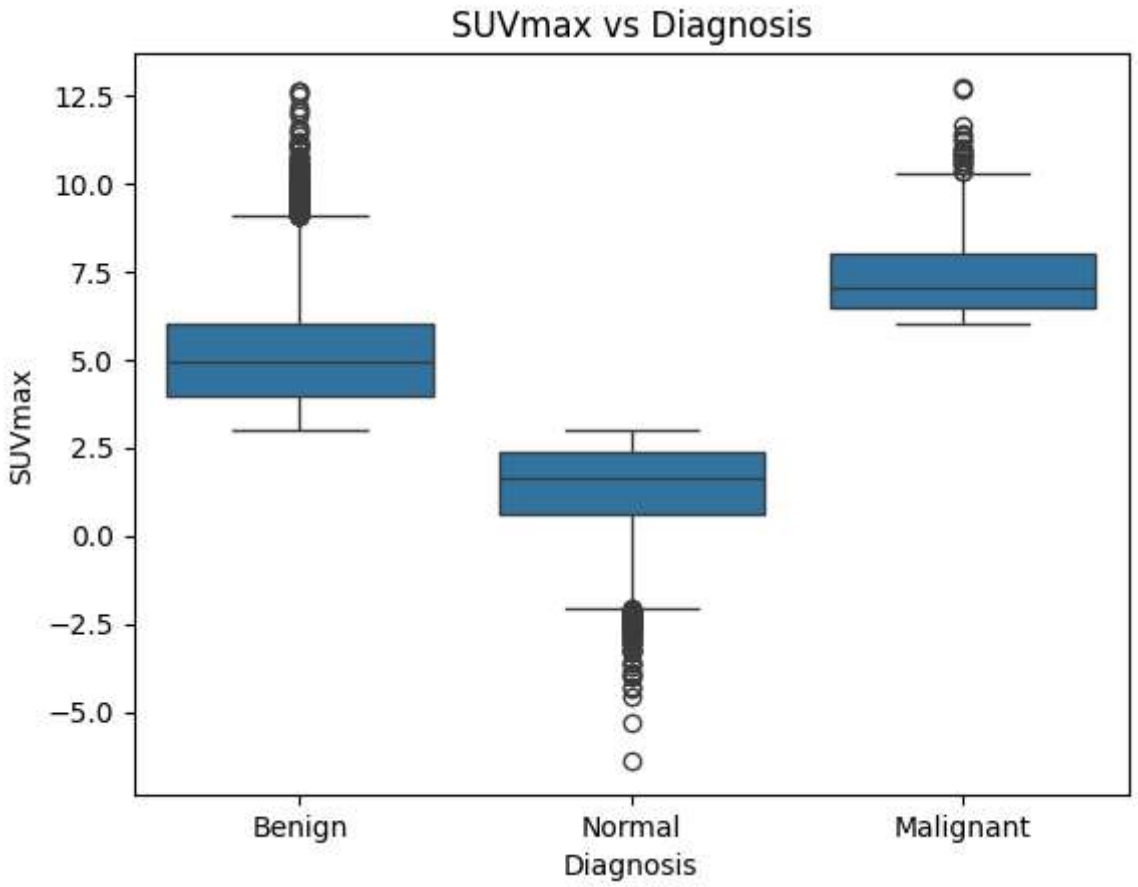
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.countplot(x="Diagnosis", data=df)
plt.title("Diagnosis Distribution")
plt.show()
```



Diagnosis Distribution Observation

Benign cases dominate the dataset, followed by normal cases, while malignant cases are least frequent. This indicates class imbalance, which is important to consider for modeling or statistical inference.

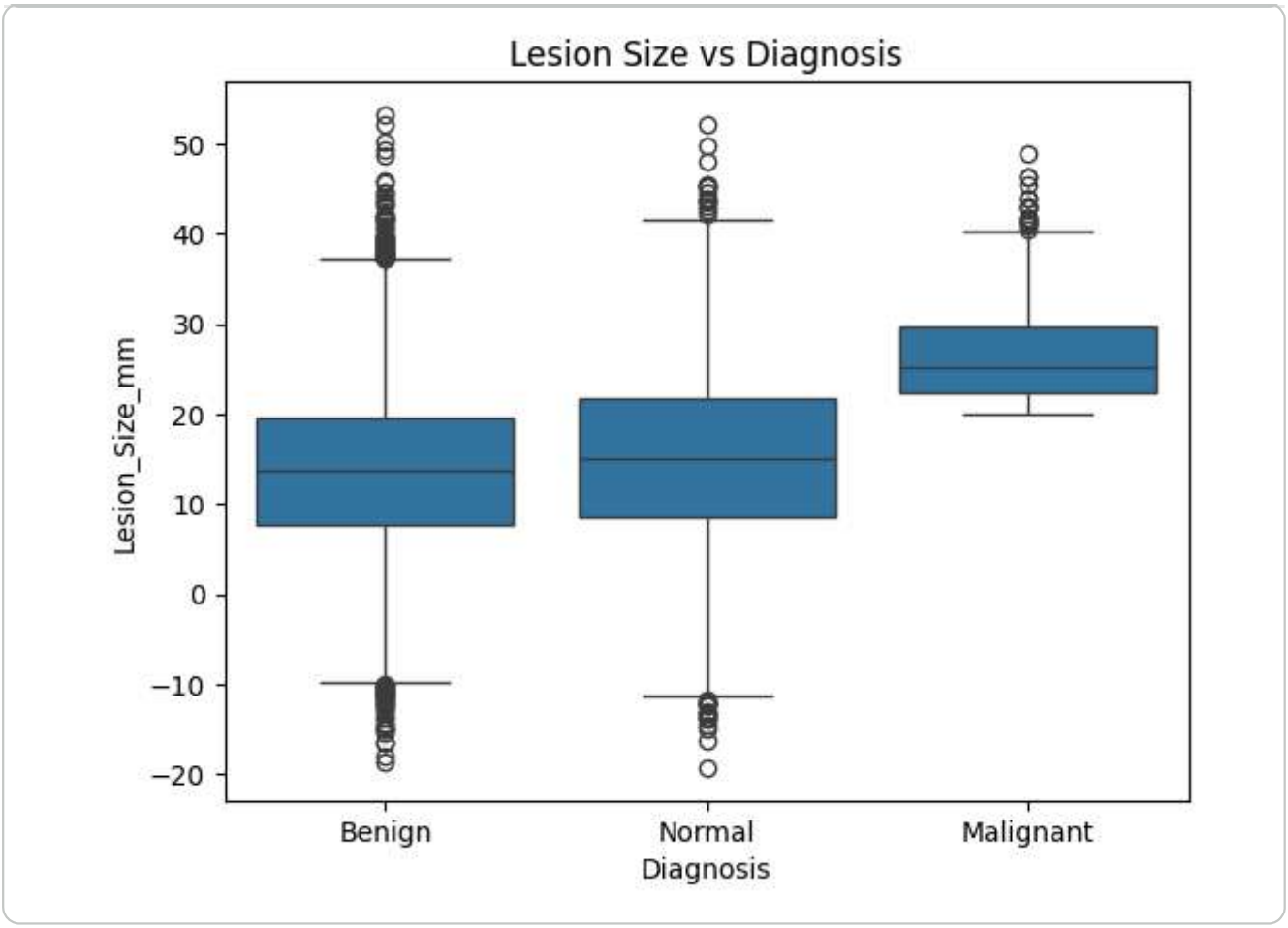
```
sns.boxplot(x="Diagnosis", y="SUVmax", data=df)
plt.title("SUVmax vs Diagnosis")
plt.show()
```



SUVmax Observation

Malignant diagnoses show higher SUVmax values on average, suggesting increased metabolic activity compared to benign and normal cases.

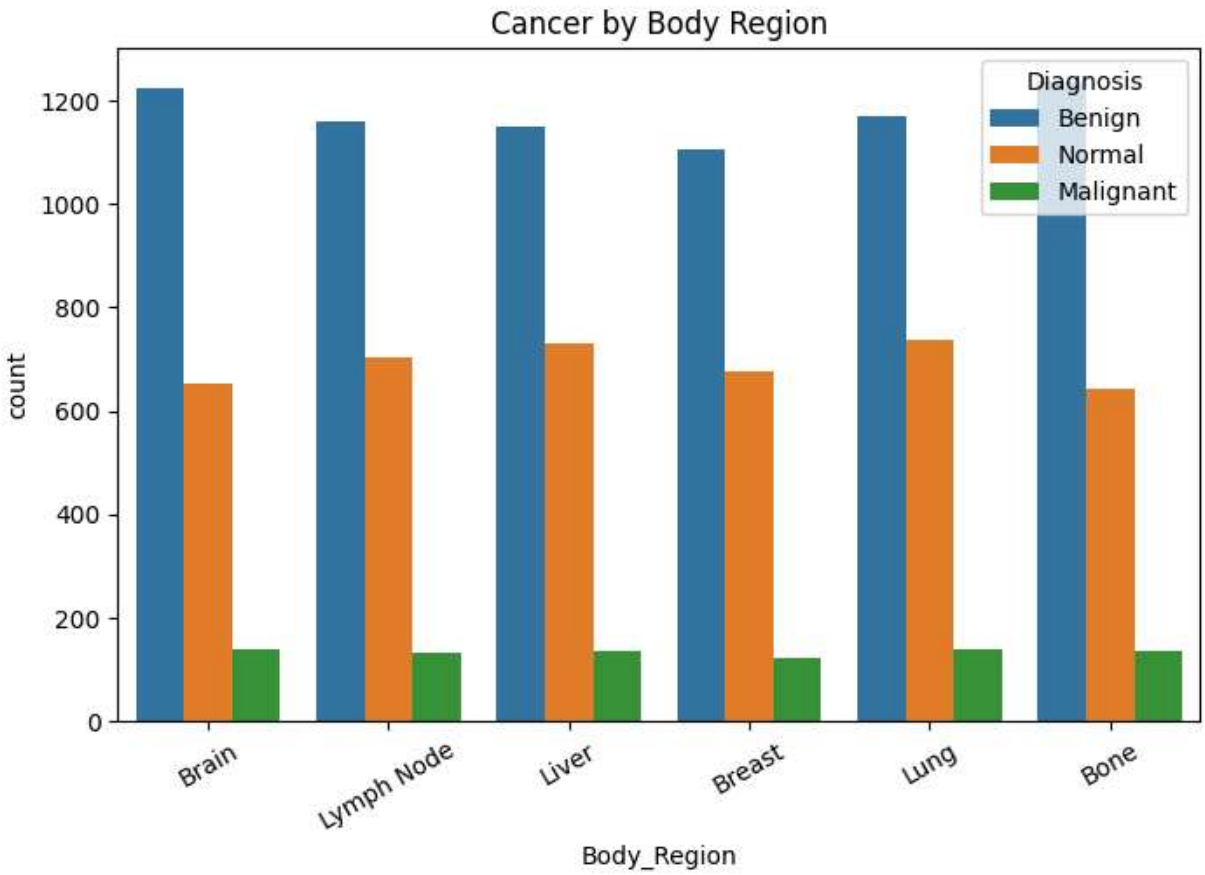
```
#Lesion Size Analysis
sns.boxplot(x="Diagnosis", y="Lesion_Size_mm", data=df)
plt.title("Lesion Size vs Diagnosis")
plt.show()
```



Lesion Size Observation

Lesion size tends to be larger in malignant cases, indicating a possible association between lesion growth and diagnosis severity.

```
plt.figure(figsize=(8,5))
sns.countplot(x="Body_Region", hue="Diagnosis", data=df)
plt.xticks(rotation=30)
plt.title("Cancer by Body Region")
plt.show()
```

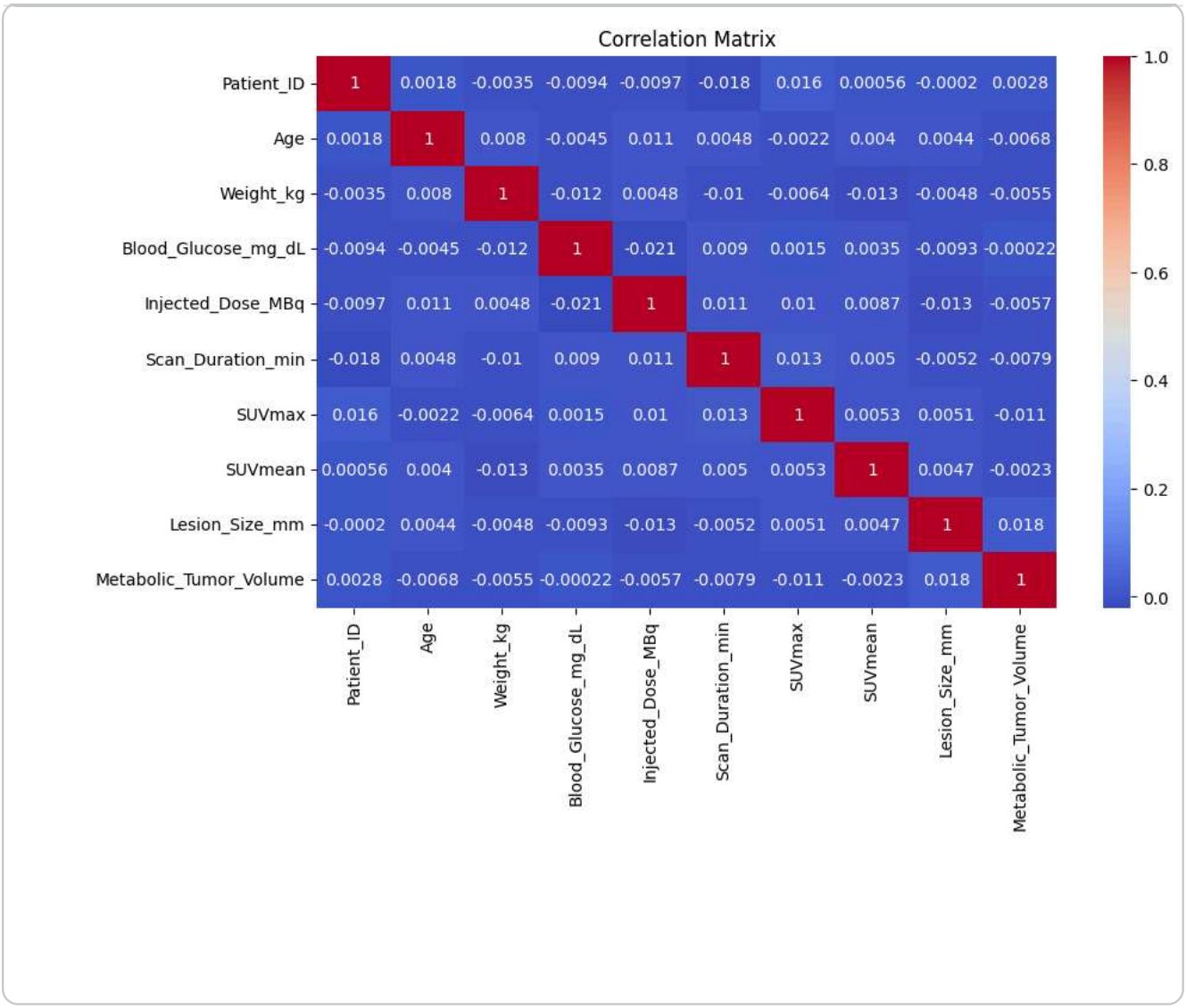


Regional Distribution Observation

Cancer cases appear across multiple body regions, with relatively consistent patterns of diagnosis distribution, suggesting no single region dominates malignant incidence in this dataset.

```
numeric_df = df.select_dtypes(include=np.number)

plt.figure(figsize=(10,6))
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Matrix")
plt.show()
```



Correlation Visual Observation

Most numerical variables show weak correlations, indicating that individual features contribute independently rather than being strongly interdependent.

Conclusion

Overall, malignant cases consistently show higher SUVmax values and larger lesion sizes, suggesting greater metabolic activity and tumor progression. Diagnosis distribution indicates class imbalance, and correlation analysis shows weak relationships among most numerical variables, implying independent feature behavior. These patterns suggest that lesion size and SUV metrics may be important predictors for classification