Differentiation of Metabolically Distinct Areas within Head and Neck Region using Dynamic ^{18}F-FDG Positron Emission Tomography Imaging

Mueez u Din

Turun yliopisto
University of Turku

Master's Thesis
University of Turku, Finland
Turku PET Centre, Turku University Hospital (TYKS), Turku, Finland
Date: July, 2013

Master's Degree Programme in Biomedical Imaging
Specializing theme: *In vivo* & Clinical Imaging

Credits: 45 ETCS

Supervisor:
Tony Shepherd
Data Analysis and Mathematical Modelling
Turku PET Centre, Turku University Hospital (TYKS), Turku, Finland

Examiners:
1: Tony Shepherd
2: Pekka Hänninen

Passed:
Grade:

As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

Albert Einstein

UNIVERSITY OF TURKU

Turku PET Centre, Turku University Hospital (TYKS),

Turku, Finland

MUEEZ U DIN

Differentiation of Metabolically Distinct Areas within Head and Neck Region using Dynamic ^{18}F-FDG Positron Emission Tomography Imaging

Master's Thesis

Specializing theme: *In vivo* & Clinical Imaging

July, 2013

Abstract:

Positron Emission Tomography (PET) using ^{18}F-FDG is playing a vital role in the diagnosis and treatment planning of cancer. However, the most widely used radiotracer, ^{18}F-FDG, is not specific for tumours and can also accumulate in inflammatory lesions as well as normal physiologically active tissues making diagnosis and treatment planning complicated for the physicians. Malignant, inflammatory and normal tissues are known to have different pathways for glucose metabolism which could possibly be evident from different characteristics of the time activity curves from a dynamic PET acquisition protocol. Therefore, we aimed to develop new image analysis methods, for PET scans of the head and neck region, which could differentiate between inflammation, tumour and normal tissues using this functional information within these radiotracer uptake areas. We developed different dynamic features from the time activity curves of voxels in these areas and compared them with the widely used static parameter, SUV, using Gaussian Mixture Model algorithm as well as K-means algorithm in order to assess their effectiveness in discriminating metabolically different areas. Moreover, we also correlated dynamic features with other clinical metrics obtained independently of PET imaging. The results show that some of the developed features can prove to be useful in differentiating tumour tissues from inflammatory regions and some dynamic features also provide positive correlations with clinical metrics. If these proposed methods are further explored then they can prove to be useful in reducing false positive tumour detections and developing real world applications for tumour diagnosis and contouring.

Keywords: Dynamic PET Imaging, Head and Neck Cancer, Machine Learning

List of Abbreviation:

ATP:	Adenosine tri-phosphate
AUC:	Area under the curve
BGL:	Blood glucose level
Bq:	Becquerel
BTV:	Biological tumour volume
ca:	Carcinoma
CT:	X-ray computed tomography
EM:	Expectation–maximization
FDG:	Fluorodeoxyglucose
FOV:	Field of view
FPR:	False positive rate
GLUT:	Glucose transporter protein
GMM:	Gaussian mixture model
HPV:	Human papilloma virus
IMRT:	Intensity modulated radiation therapy
LBS:	Lutetium based scintillators
MRI:	Magnetic resonance imaging
NPU:	Normal physiological uptake
PET:	Positron emission tomography
ROC:	Receiver operating characteristics
ROI:	Region of interest
RI:	Retention index
SCC:	Squamous cell carcinoma
SUV:	Standardised uptake value
TACs:	Time activity curves
VOI:	Volume of interest

Table of Contents:

1. Introduction

Medical imaging is the field of acquiring and analysing images of internal structures and functions of the body that can aid in diagnosis of a disease and also assist in its treatment planning (LeVine, 2010). Out of various available medical imaging techniques, positron emission tomography (PET) imaging with the ^{18}F-FDG radiotracer has been used for decades for the diagnosis of cancer and its utilization in the process of radiotherapy planning is also increasing with the passage of time. Several studies have demonstrated the effectiveness of PET in intensity modulated radiation therapy (IMRT) planning and defining biological tumour volume (BTV) (Minn et al., 2010). The radiotracer used in PET imaging accumulates in different areas of the body depending upon its intrinsic properties and acquired PET images represent these areas of higher radiotracer accumulation as regions of higher intensity or 'hot-spots'. However, the most widely used radiotracer, ^{18}F-FDG – an analogue for glucose, is not specific for tumour tissues and it can also accumulate in benign inflammatory or infectious lesions, similar to that of tumours, because of higher glucose metabolism in those particular areas (Bahkeet and Powe, 1998). This limitation of ^{18}F-FDG radiotracer makes the image interpretation challenging particularly in the imaging of head & neck region where, quite commonly, tumour and inflammation reside close to each other and differentiation of areas of higher uptake into malignant, benign, inflammation and normal tissue becomes crucial for the optimum diagnosis and radiotherapy planning (Yao and Faulhaber, 2012).

In PET, activity concentration of the radiopharmaceutical is first measured in units of Becquerels per millilitre (Bq/ml). For imaging, particularly in oncologic PET, this unit is commonly converted to standardised uptake values (SUVs). However, SUV calculation introduces other limitations as well as those of the tracer, regarding the adequate differentiation of functionally distinct metabolic areas. SUV is calculated by aggregating the PET signal over a selected time window, typically several minutes, and afterwards normalising it with respect to patient's body mass and initial concentration of the injected radiotracer. As well as losing dynamic information in the temporal aggregation, SUV also depends upon many other confounding factors which include

PET frame duration, patient's diet, scanning protocol and image reconstruction (Thie, 2004; Boellaard, 2009). While basing judgement for tissue discrimination exclusively on SUV, dynamics of the radiotracer accumulation in the tissues are being neglected as it is relying primarily on the averaged concentration over time.

The reason for the elevated glucose or ^{18}F-FDG radiotracer metabolism in the tumour and inflammation is multifactorial; some of these factors that have been mentioned by Gillies et al. (2008) and Plathow et al. (2008) are due to the up-regulation of glucose transporters (GLUTs) and increased enzymatic activity of hexokinase proteins. The up-regulation of the GLUTs leads to enhanced tumour glucose metabolism (Younes et al., 1997). Increased level and activity of the hexokinase enzymes is also another reason for increased uptake of ^{18}F-FDG in the tumour cells. Jadvar et al. (2009) have postulated that there can be a different level and extent of GLUT and hexokinase expression in normal tissue, cancer and inflammatory regions which can be evident in the different temporal profile of ^{18}F-FDG uptake in these different areas. On the basis of that acquired information the areas of FDG uptake, which may be similar looking in a static time averaged PET image, can be differentiated based on their temporal profile.

Dynamic PET acquisitions are gaining interest for radiotherapy nowadays because radiotracer metabolism in the area of interest can be followed with the time profile (Muzi et al., 2012; Shepherd and Owenius, 2012a). If we assume that there will be a different extent and level of GLUT and hexokinase expression in the tumour, inflamed regions and normal tissue surrounding tumours, then we can infer that the metabolic pathways in these areas will be different and with the aid of a dynamic PET acquisition protocol, a map of ^{18}F-FDG radiotracer dynamics can be used to differentiate metabolically different areas.

In summary, current imaging and image processing methods are limited for differentiating functional volumes of interest in FDG PET, and recent evidence suggests that dynamic PET can add relevant, discriminant information. This thesis presents a new dynamic imaging protocol for PET oncology, then develops and evaluates analysis techniques for the dynamic image data that aim to improve on SUV imaging for diagnosis and treatment planning.

2. Review of Literature

2.1 Positron Emission Tomography

Positron Emission Tomography, commonly known as PET, is a non-invasive imaging modality which uses radiopharmaceutical drugs to image molecular function. After injection of a radiopharmaceutical drug, it circulates throughout the entire body as a normal physiological mechanism and at the same time it starts accumulating at specific sites depending upon its binding properties (Saha, 2010). During the whole process the radiopharmaceutical keeps on emitting positrons as a process of nuclear decay. The reason for the process of nuclear decay, in these molecules, is a higher number of protons which makes nuclei unstable; as a result, a proton decays into a positron, a neutron and a neutrino. A positron is an antiparticle having a positive charge and mass equal to that of an electron; after emission, positrons travel some distance and collide with their antiparticles, electrons, by a process known as annihilation. The electrons involved in annihilation can belong to any molecular structure of any local tissue. The distance travelled by a positron before colliding with an electron depends upon energy with which it is emitted and density (electron) of the surrounding tissue. Annihilation converts mass m of these particles into energy e, according to the equation $e=mc^2$, where c denotes the speed of light. This conversion produces energy in the form of two gamma photons travelling in opposite directions to each other (almost 180^o) from the point of annihilation and having energy of 511 KeV. These emitted gamma photons are detected by the PET camera detectors which are designed in such a way that they surround the patient's body in a form of a ring (Figure. 1 & Figure 2A). Imaging is based on the assumption that any pair of photons that arrive opposite to each other, and at the same time, originate from the same annihilation event. The device estimates the 'line of response' between the two detectors corresponding to one annihilation, and the actual point of annihilation along that line is estimated for all annihilations in a single time frame by a reconstruction algorithm, which also uses the time difference between

paired photons reaching the detectors. All the acquired raw data are sent to the computers for tomographic image reconstruction. (Cherry, 2006.)

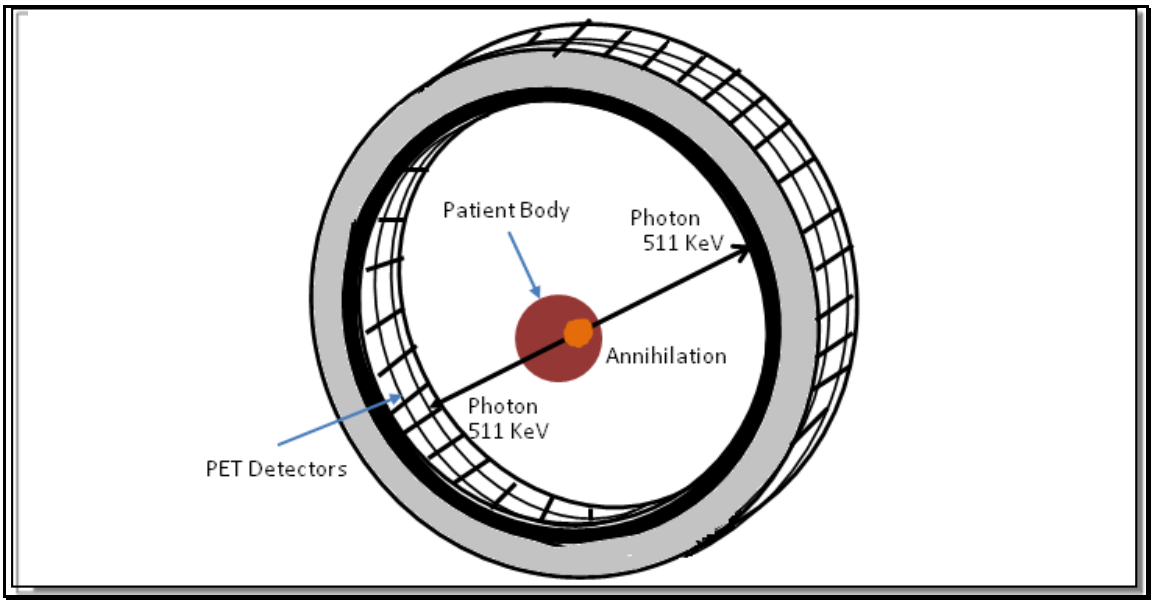


Figure 1: *Diagrammatic representation of the PET detectors designed such a way that they surround the patient's body. Gamma photons, produced as a result of annihilation, are detected in coincidence by the PET detectors and it is marked as one single event.*

Figure 2 shows the scanner used for all PET imaging in this study. It is a hybrid PET-CT device that also acquires X-ray CT images of the patient for attenuation correction and for anatomical reference by fusing with PET. The device also has time of flight technology to improve the PET resolution; with the aid of this technology the location of the annihilation is also estimated by the time difference (of the order nanoseconds) between two photons reaching detectors.

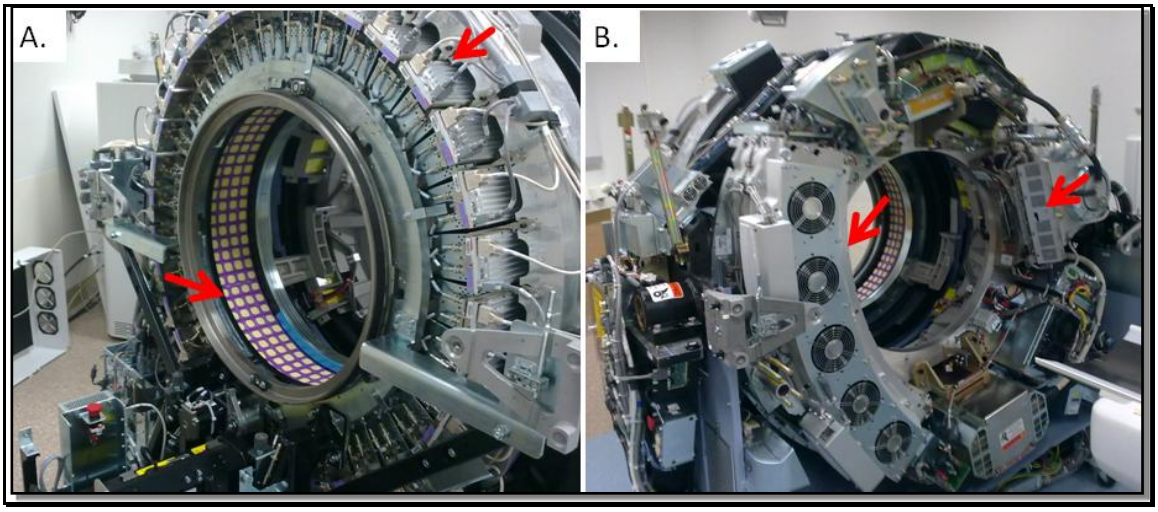


Figure 2: *Photographs of the opened up gantry of PET-CT scanner (General Electric Discovery 690 PET-CT Scanner) A. PET assembly of the scanner; red arrows point to PET detectors (bottom left) and photo multiplier tubes (top right). B. CT assembly of the scanner; red arrows point to x-ray tube (right) and x-ray detectors (left). Image courtesy: Tony Shepherd, Turku PET Centre, Turku University Hospital (TYKS), Turku, Finland.*

2.1.1 ^{18}F-FDG – Fluorodeoxyglucose

^{18}F-FDG, which is an analogue for glucose and has a half-life of 110 minutes, is the most commonly used radiotracer in the PET imaging (Delbeke et al., 2002). In particular, it is very useful in oncological PET examinations because of the observation that rapidly dividing tumour cells require higher energy in the form of adenosine tri-phosphate (ATP) and cells generally fulfil their energy requirement by the process of breaking down glucose known as glycolysis (Miles and Williams, 2008). In order to have higher energy, cells need to take in higher amounts of glucose and during PET examinations ^{18}F-FDG, being a glucose analogue, accumulates at the area of higher glucose utilization. Over time as the tracer washes away from tissue regions that do not capture FDG in this way, the regions of FDG utilization become most visible in PET

images. A more detailed description of the mechanism of glucose and ^{18}F-FDG metabolism in tumours, normal and inflammatory tissues is given in section 2.3.

2.1.2 Imaging Protocols

PET scans can be acquired with two main types of acquisition protocol, static and dynamic. Static scanning acquisition protocols are the most commonly employed nowadays when imaging for radiotherapy treatment planning. In these protocols the scanning is started a certain period of time after radiotracer injection and the activity concentration of the radiotracer is measured over a fixed period of time, around 2 to 5 minutes. For the static scan, with ^{18}F-FDG radiotracer, typical clinical practice is to start image acquisition after approximately 60 minutes of radiotracer administration (Boellaard et al., 2010).

With the aid of dynamic PET acquisition protocol, activity of the radiotracer can be measured at multiple time points and this activity concentration plotted against time to form time activity curves (TACs). A TAC provides information about the kinetics of the radiotracer in the specific tissue and at the same time gives insight about the rate of tracer in and out flux and amount/rate of tracer accumulation (Thorwarth et al., 2005).

The whole body scan of a typical adult cannot be done in a single bed position because the field of view (FOV) governed by the assembly of the PET detectors is not big enough to cover all the body regions at the same time. Whole-body PET acquisitions are therefore made from a series of multiple bed positions. At each bed position, the scanner acquires data for the set period of time and then moves along in the axial direction to the next position.

2.1.3 Image Analysis

After PET acquisitions and image reconstructions have been done, a 3D image volume is produced for the image analysis purposes. Inside an image-volume, each individual voxel (volume cell) represents the activity concentration in the corresponding anatomical tissue (Cherry, 2006). Size of the voxel depends upon the image reconstruction algorithm employed. In a typical image, having 3.27mm slice thickness, in-slice field of view 70 x 70 cm and matrix grid-size 192 x 192, each of the voxels represent 0.0435 cm^3 (3.65 x 3.65 x 3.27mm) of the corresponding anatomical tissue. On a normal image display the radioactivity concentration is represented as a certain value of grey scale (or pseudo colour). In clinical practice this activity concentration is normalised with respect to the injected radioactive concentration and patients' body mass and commonly referred to as standardised uptake value (SUV). The SUV is calculated using the following equation,

$$SUV = \frac{Activity\ concentration}{Injected\ Activity\ /\ Body\ Mass}$$

(Equation 1)

This measurement of SUV is treated as an index for tissue metabolism as well as degree of malignancy (Nahmias and Wahl, 2008). Figure 3 shows typical PET images (fused with CT images for anatomical reference) of the head and neck region obtained using ^{18}F-FDG radiotracer. The intensity of each PET voxel is in units of SUV.

It has been documented that SUV is a semi quantitative value and has several pitfalls; various factors can produce inconsistent results for the same volume of interest. These factors include patient's body size, time of image acquisition after injection, patient's plasma glucose level, and possible dose extravasations (Lin and Alavi, 2009). Laffon et al. (2011) proposed a method of adjusting SUV for injection-acquisition time

differences, but the method has limited practicality in the current clinical setting and until now is only validated in cases of lung cancer.

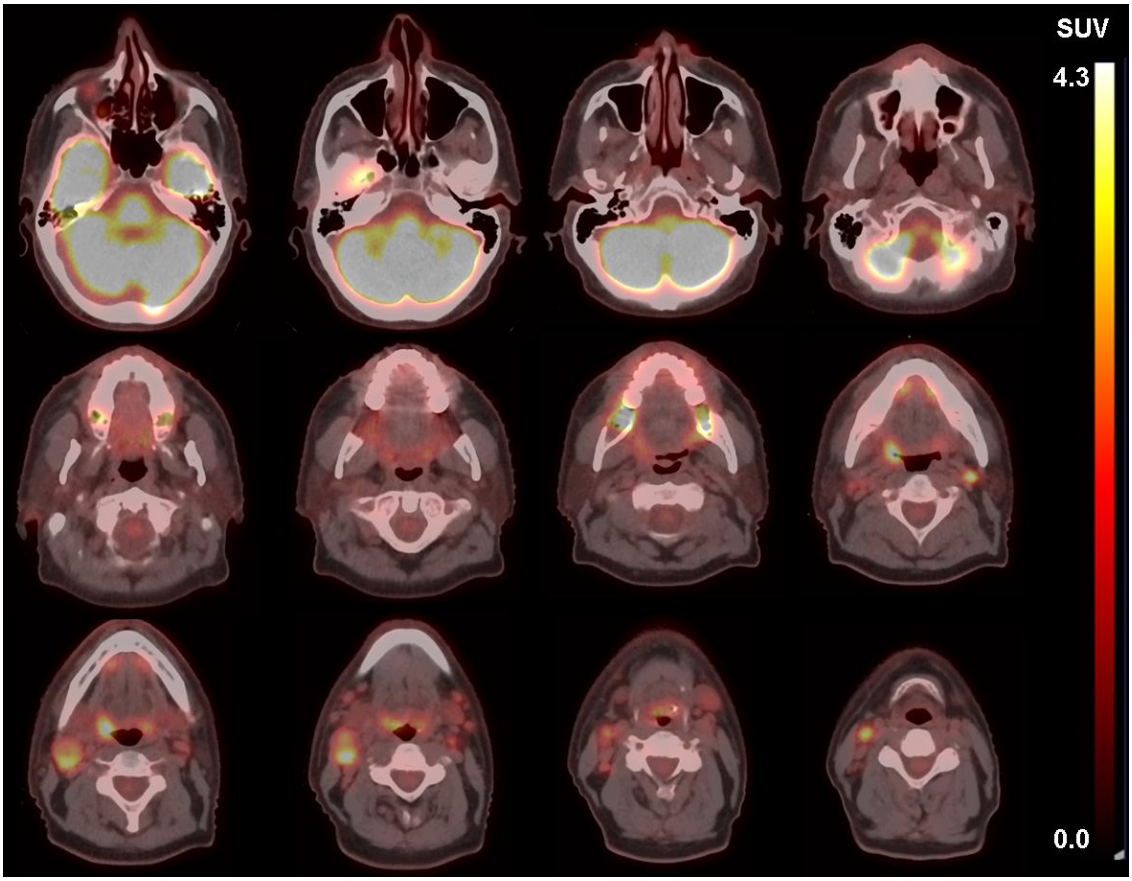


Figure 3: *Axial sections of fused PET and CT images of head and neck region; the patient has palatine tonsil carcinoma.* ^{18}*F-FDG radioactivity concentration has been shown on the SUV scale. Multiple uptake areas represent higher glucose metabolism within those regions.*

Moreover, while imaging with ^{18}F-FDG radiopharmaceutical, thyroid glands, lymph nodes, myocardium, site of infection, skeletal muscle, and benign pathological FDG uptake in healing bones can represent similar SUV to that of cancer and thus cause interpretive pitfalls (Shreve et al., 1999).

Several methods have been suggested in order to differentiate similar looking uptake areas on the PET images. Some of the studies have suggested using dual time point PET

acquisition protocol; Zuhang et al. (2001) using dual time point imaging showed that the SUV of benign lung nodules slightly decreased, while average SUV of the malignant lesion increased in the second time point. Similarly, Chui et al. (2012) reported a successful discriminatory technique using dual time PET acquisition protocol for differentiating malignant solid lung nodules from benign ones. Basu et al. (2009) have shown similar phenomena over an extended period of time (upto 8 hours) in patients of lung carcinoma; Lodge et al. (1999) observed that FDG uptake in high grade sarcomas continued to increase for up to four hours whereas FDG uptake came to an equilibrium state within thirty minutes in normal tissues; however, any particular imaging analysis technique or imaging protocol has not been suggested, leaving the issue unresolved. Hustinx et al. (1999) also utilised a dual time point imaging protocol in head and neck cancer patients and suggested that if the two scanning sessions are 30 minutes apart then the tumour uptake of FDG may be distinguishable from those of normal and inflamed tissues based on the change in activity concentration.

By using full dynamic imaging of several time points, compartmental modelling has also been used in various researches for the quantification of radiotracer kinetics within the tissue of interest (Morris et al., 2004). However, compartmental modelling is limited for applications in radiotherapy treatment planning. One limitation is the arterial input function that needs to be provided; this function is derived from the radioactivity concentration within the blood supply, measured at various time points. Measurement of radioactivity concentration within blood supply require repetitive blood sampling during the period of scanning which increases patient discomfort as well as work burden for the operators. Another limitation is that models are sensitive to noise and are therefore calculated for spatially averaged TACs within pre-defined volumes of interest.

2.2 Head and Neck Carcinoma

In head and neck malignancies the most common type of malignant tumour is the one that originates from the surface of the squamous epithelium and it is commonly known as Squamous cell carcinoma (SCC) (Wenig and Cohen, 2009). SCC of head and neck

region is newly diagnosed for over half a million patients annually worldwide (Minn et al., 2010) and it also represents 2-4% of all the malignant conditions in United States (Carvalho et al., 2005).

Cigarette smoking and alcohol consumption are the most common causes of the development of head and neck malignancies (Lewin et al., 1998). Moreover, Human papillomavirus (HPV) has been reported to be associated with SCC (mainly oropharyngeal) of head and neck regions (Syrjänen and Syrjänen, 2000).

SCC can be further classified into different grades based on histopathologic characteristics. Under high magnification, the more tumour resembles the appearance of a cell and/or tissue layer from which it has arisen, the better differentiated it is. If seen using microscope with proper staining, a well differentiated SCC has obvious keratinisation and intercellular bridges, and it is given grade I; moderately differentiate carcinoma is labelled as grade II, poorly differentiated as grade III and undifferentiated as grade IV. (Wenig and Cohen, 2009.)

As part of the diagnosis in the case of head and neck cancer, physicians also assign a certain stage (commonly known as TNM staging) based upon anatomic extent of the primary tumour (T), the involvement of regional lymph nodes (N) and distant metastasis (M) (Sobin, 2003).

Syrjänen (2004) has mentioned that tonsils, composed of many crypts, are mainly populated by lymphocytes and their surface is covered by stratified squamous epithelium, while a similar surface also lines the crypts. It has been believed that HPV associated tonsillar SCC originates from the epithelium of the crypts, while non-HPV associated SCC arise from tonsillar surface epithelium (Frisch et al., 2000 and Hemminki et al., 2000).

2.3 Mechanism of Glucose / ^{18}F-FDG Metabolism

Glucose is metabolised in tissues with the help of the glucose transport protein: GLUT. GLUT is involved in the first rate-limiting step for glucose metabolism that allows energy independent glucose transport across the cell membrane down the concentration gradient (Macheda et al., 2005). The mechanism of the metabolism of glucose, as well as the glucose analogue ^{18}F-FDG used in PET imaging, is described in detail below for different tissues i.e. normal tissue, head and neck tumours and inflammatory tissues.

2.3.1 Normal tissue

Sugar, mainly glucose, is one of the major sources of energy to a normal animal cell. For cells to take in polar molecules like glucose, the molecules need to cross a semi-permeable plasma membrane with the aid of membrane-associated carrier proteins. GLUT is one of the families of these membrane-associated glucose transporter proteins which use the gradient of sugar concentration inside and outside the plasma membrane to facilitate glucose movement through that membrane (Calvo et al., 2010). Following entry of the glucose molecules into the normal cells, they are converted to pyruvate as a result of glycolysis. Pyruvates subsequently change into acetyle-CoA which is afterwards used by mitochondria to generate ATPs (Szablewski, 2013). Hexokinase enzyme binds to the mitochondrial membrane and phosphorylates glucose to glucose-6-phosphate using the energy produced by the mitochondria. The end product of this glucose metabolism in a normal cell, under aerobic conditions, is carbon dioxide and water. Similarly, ^{18}F-FDG is phosphorylated to FDG-6-phosphate, but contrary to glucose-6-phosphate, it cannot be metabolised further in the glycolytic pathway and becomes trapped in the cell because of its negative charge.

2.3.2 Head and Neck tumours

Basic glucose metabolism reactions in tumour cells are similar to the normal tissues mechanisms; however, in case of FDG metabolism in the tumours cells, the low activity

of the reverse enzyme, glucose-6-phosphotase, leads to the tumour cell accumulation of FDG-6-phosphate which is visible in the PET images as bright spots because of the abundant annihilation reactions at the corresponding location of the patient's body and by the complex process of image formation described earlier in this document in section 2.1.

On the basis of sequence similarity GLUT has been divided into three classes (Class I, II & III) and each class is further subdivided into different subcategories. Class I consists of GLUT-1, GLUT-2, GLUT-3, GLUT-4; Class II consist of GLUT-5, GLUT-7, GLUT-9, GLUT-11 and Class III consist of GLUT-6, GLUT-8, GLUT-10, GLUT-12. It has also been reported that the glucose transport proteins involved in the transport of glucose across the tumour cells' membrane express themselves in tissue-specific fashion. GLUT-1 and GLUT-3 have been reported to be involved in the basal glucose uptake of extra-cranial head and neck tumours (Mellanen et al., 1994) while GLUT-2 & GLUT-4 have not been detected. Several studies have also mentioned that GLUT-1 and GLUT-3 gene expression have been found to be higher within tumour cells of head and neck cancer as compared to surrounding normal tissues (Zhou et al., 2008).

2.3.3 Inflammation

Essentially, glucose transport mechanism is the same in inflammatory tissues as it is in normal tissues; however, it has been reported that the expression of glucose transporter proteins may be different in inflammatory tissues as compared to normal and cancerous tissues. White blood cells and platelets are abundant in inflammatory tissues (Bailey et al., 2006) and GLUT-3 has been reportedly detected in platelets, as well as in the entire set of white blood cells (lymphocytes, neutrophils, monocytes and macrophages) (Calvo et al., 2010). GLUT-3 has higher affinity for the transport of glucose molecules compared to GLUT-1 (Simpson et al., 2009).

As it has been documented in the published literature, there is difference in the mechanism of glucose metabolism in tumours, compared to other metabolising tissues.

It is reasonable to expect that this difference can be observed in the TACs of different areas using ^{18}F-FDG radiotracer, and that tissue discrimination can be based on those differences in TACs.

2.4 ^{18}F-FDG PET Imaging of Head and Neck Carcinomas

PET imaging fused with CT has been reported to be of high diagnostic performance in the assessment of head and neck cancers (Gordin et al., 2007). ^{18}F-FDG PET alone has been found to be more accurate in detecting primary as well as metastatic disease in head and neck region as compared to CT or MRI (Adam et al., 1998; Kresnik et al., 2001; Haerle et al., 2011); however, PET offers limited anatomical demarcation; therefore, it is combined with another imaging modality, commonly CT or MRI for tumour delineation as well as for precise anatomical differentiation. Moreover, PET is also reported to be a highly useful imaging modality in detecting any residual or recurrent tumour in head & neck after radiation therapy or surgery (Hubner et al., 2000).

Despite the high sensitivity and specificity of PET in detecting tumours of the head and neck region, there are still problems with the delineation of the gross tumour volume (GTV) with PET imaging for radiotherapy planning. As well as limitations of the imaging technology such as low resolution and partial volume effects (Zaidi et al., 2009) the processes governing the tracer uptake and washout over the period of time are complex, and not clear in SUV images. For example, more than one process such as metabolism and hypoxia can act together, and inflammation may be present to some extent within a tumour region, making static ^{18}F-FDG PET alone insufficient for discriminating between different uptake areas in terms of tissue function.

2.5 Classification

In general terms, classification is a method of identifying a category for a new observation on the basis of data that have already been categorised based on some chosen features (Pham and Bazin, 2008). Classification is a fundamental tool in medical imaging. In screening for breast cancer for example, analysis of the whole mammograph may be used to classify a patient as normal or suspected of disease. Furthermore, within diseased patients, tumours can be further classified, e.g. Guo et al. (2009) showed that characterisation and classification of tumour lesions is possible using computerized fractal-based texture analysis in digital mammograms.

Feature selection and extraction are important steps used in classification of medical images, which means selecting the subset of features that are more likely to distinguish one tissue class or patient diagnosis from another (Hall et al., 1971). As in this thesis, voxels may be labelled as belonging to a functional or anatomical class based on the chosen features.

In order to first identify features on which classification can be based, clustering can be used; where clustering is a technique of grouping data in such a way that the data points in the same cluster are more similar to each other than to other data points in other cluster, based on the chosen features (Ng et al., 2006). There are several clustering algorithms proposed for the classification of data; however, discussing each of them is beyond the scope of this thesis document. Generally, classification algorithms have been divided into two main classes, parametric and non-parametric, and for the purpose of classification of our data we choose to use one common classification algorithm from each of these main classes. Another distinction can be made between supervised and non-supervised classification algorithms. Supervised methods use labelled training data to create inferred functions so that new observations from outside the training set can be assigned to the already defined classes using training data. Unsupervised classification algorithms do not require training data; rather, they divide observations into a (usually pre-defined) number of classes.

In our present study we consider one PET voxel as one observation or data point so that the task becomes classifying image regions according to their tissue function. We also decided to investigate those properties on the basis of which we can cluster data points of the same volume of interest (VOI) together and differentiate them somehow from other VOIs. These features or properties were investigated and derived from the time activity curves (TACs); these are further discussed in detail in section 4.4. We aimed to use two clustering algorithms, Gaussian Mixture Model (GMM) clustering algorithm and K-means clustering algorithms; these algorithms have been discussed one by one in detail below.

2.5.1 The Gaussian Mixture Model

Gaussian Mixture Model is a parametric method of unsupervised clustering in which there are no predefined classes and the algorithm assumes all the data points to be normally or Gaussian distributed. The complete Gaussian mixture model is parameterized by the mean vectors, covariance matrices and mixture weights from all component densities (Ion, 2011).

The Gaussian mixture model can be generally represented by the following equation:

$$f(x, \mu, \Sigma, w) = \sum_{i=1}^{m} w_i g(x, \mu_i, \Sigma_i)$$

(Equation 2)

Where x is the D-dimensional continuous-valued feature vector with mean vector μ and covariance matrix Σ and w_i $(i = 1 \ldots m)$ are the mixture weights from all component densities; m is the number of classes and in the case of binary classification m is equal to 2 (Reynold, n.d). GMM parameters are estimated from a given dataset using an optimization algorithm, for example the iterative Expectation-Maximization (EM) algorithm; initially, assumptions are made about parameters μ and $\sum$, and then the estimates are iteratively refined in order to maximize the likelihood of the observed data

according to those parameters. The advantage of using EM algorithm is its capability for handling uncertainties within a given dataset (Dempster et al., 1977; Bishop, 2006).

GMM has been used previously in various medical image segmentation techniques; Górriz et al., (2009) has used GMM in three dimensional functional brain images to divide images into clusters according to region of low and high activation using the intensity distribution of the voxels. The resulting ROIs are claimed to be useful in computer aided diagnosis of brain diseases like Alzheimer's disease.

In our study we also aimed to differentiate different VOIs using this clustering algorithm based on feature vectors obtained from TACs of the different VOIs. GMM assumes the feature vector to be Gaussian distributed; however, some of the feature vectors obtained from the TACs might not be normally distributed thus causing difficulty for the GMM algorithm to differentiate between the two different classes. In this situation K-means algorithm can in principle demonstrate better results as compared to GMM algorithm.

2.5.2 K-means Clustering Algorithm

The K-means clustering algorithm is an unsupervised learning algorithm belonging to the non-parametric class of clustering algorithms. It follows a simple way to classify a given data into a predefined number of clusters (k) (Pham et al., 2000). The algorithm works such that each of the data point's membership is determined by approximating the centroid for each cluster and assigning each data point to the cluster with the nearest centroid (Su and Chou, 2001); for minimizing the cluster point dispersion the algorithm iteratively relocates the centroid point of each of the k clusters (Oyelade et al., 2010).

This clustering algorithm is widely being used for medical image segmentation in several areas of research; however, the use of this algorithm for the differentiation of different metabolic areas in PET images is limited. Jansen et al. (2009) utilised a K-means algorithm to classify PET images based on a single feature: the slope of the time

activity curves obtained from PET images of the rectal cancer patients. The method distinguished tumour from healthy tissue and provides motivation for further use of dynamic PET classification such as in this thesis.

2.5.3 Classifier Performance Evaluation

In order to measure the performance achieved by our chosen classification algorithms we used different performance evaluation techniques. One of the ways for evaluating the success of a classification algorithm is receiver operating characteristic (ROC) analysis. A ROC curve is a plot of the sensitivity of an algorithm's classification versus 1-specificity or false positive rate (FPR) for different thresholds of a controlling parameter (Park et al., 2004; Murphy, 2007). The area under the curve (AUC) produced by the ROC analysis can be used as a combined measure to evaluate the sensitivity and specificity of the performance of the classifier algorithm (Metz, 2006) and it is closely related to Wilcoxon rank-sum test (Hanley and McNeil, 1982; Mason and Graham, 2002). For the performance evaluation of our GMM we used AUC of the ROC curve; whereas, in order to measure success of K-means clustering algorithm we directly used the sensitivity and specificity of the success of the classifier. In a classification experiment, sensitivity may refer to the proportion of the true positives correctly identified as such; whereas, specificity may refer to the proportion of the negatives which are correctly identified (Altman and Bland, 1994).

All of the above performance measures require some definition of 'truth' used in counting true/false positive and negative classifications. In general, test data should be independently labelled as belonging to one class or the other. In medical imaging, without using simulated or phantom data, the best available approximation of truth is used and is never perfect (Shepherd et al., 2012b). The gold-standard method is to use manual delineation, revisited in section 4.3.1.

3. Aims and Hypotheses

3.1 Aims

The aims of this master's thesis were to present a new dynamic PET imaging acquisition protocol for head and neck cancer imaging and radiotherapy planning and to devise new image analysis techniques for tissue discrimination, with particular reference to contouring for treatment planning. Moreover, it also intended to compare the new analysis methods with standard methods in terms of sensitivity and specificity and to evaluate the power of new imaging parameters (features) for diagnosis and staging.

3.2 Hypotheses

The thesis work was based on the following broad hypothesis:

- Features derived from time activity curves (TACs) using dynamic PET can differentiate disease from inflammation and healthy tissue with significantly higher sensitivity and specificity than currently used, static parameters.

Later on, specific hypotheses were tailored according to early results and the nature of arising test data; these hypotheses are:

$\mathcal{H}$1: Uptake areas of inflammation can be better differentiated from tumour areas using dynamic features, derived from TACs, compared to SUV.

$\mathcal{H}$2: Healthy tissue surrounding tumour can be better differentiated from tumour tissue using features, derived from TACs, compared to SUV.

$\mathcal{H}$3: Normal physiological uptake areas can be differentiated from tumour's uptake regions using features derived from TACs.

$\mathcal{H}$4: Mean dynamic feature (extracted from TACs) of tumour VOIs better discriminate between tumours of different stages compared to SUV.

$\mathcal{H}$5: Mean dynamic feature (extracted from TACs) of tumour VOIs better discriminate between tumours of different grades compared to SUV.

$\mathcal{H}$6: Mean dynamic feature (extracted from TACs) of tumour VOIs better discriminate between tumours of different HPV status *(HPV-positive or HPV-negative)* compared to SUV.

4. Materials and Methods

4.1 Patients

In total five patients (age 58 to 65 years; all males) were included in the study, all having squamous cell carcinoma (SCC) of head and neck region. Patients were numbered one to five and in order to maintain patient confidentiality, patients' ID, name and other personal information were removed from the meta-data of those images that were intended to be used for research purposes. Patients 1 – 3 had palatine tonsils carcinoma whereas patient 4 had carcinoma of the base of tongue and patient 5 had oropharyngeal carcinoma. Detailed patients' information is given in table 1.

Table 1: *Clinical and other data for the patients enrolled in the study*

Patient No.	Patient Age *(Years)*	Tumour site	Staging			Grade*	HPV Status	Injected Activity (MBq)
			T	N	M			
1	63	Tonsillar ca.	2	2B	0	III	Positive	320.39
2	65	Tonsillar ca.	3	2B	0	II	Positive	309.04
3	59	Tonsillar ca.	2	2B	0	II	Positive	298.33
4	59	Base of tongue ca.	3	2B	0	II	Positive	293.77
5	58	Oropharyngeal ca	3	2B	0	I	Negative	283.87

** Histopathologic Squamous Cell Carcinoma Grade*

4.2 Scanning Protocol

Patients, with known head and neck cancer, were scanned in treatment planning position with the GE Discovery 690 hybrid PET-CT scanner (CT: 64-slices; PET: LBS crystals, 47 transverse slices of 3.27 mm slice thickness, axial FOV 15 cm) at Turku PET Centre, Turku University Hospital, Finland using ^{18}F-FDG as radiotracer. All patients fasted for at least 4 hours before the scan. None of the patients had received any treatment intervention (radiotherapy, chemotherapy or surgery) before the PET imaging. Patients were positioned supine on a flat scanner table and immobilised with the help of restraining devices including thermoplastic face mask specifically moulded to each patient. For the delivery of the radiotracer, catheter lines were placed in the patient's body to provide the intravenous vascular access.

Figure 4 diagrammatically represent the whole scanning protocol along with time. A CT scan of the body (from the level of eyebrow to that of thorax) was acquired for the planning of PET scanning and afterwards a bolus of radiopharmaceutical drug (^{18}F-FDG) was administered and dynamic PET scan was initiated instantaneously; acquisitions were performed for up to 30 minutes and the dynamic data were reconstructed in 20 time frames. After this first, dynamic acquisition session the patients rested outside the scanner before a second session began.

After 70 minutes of the FDG injection, a routine PET-CT scan (from the level of the eye brow to that of the kidney) was performed in the second imaging session comprising static acquisitions that covered the whole body in four bed positions. The first two bed positions were imaged with four minutes acquisition time, and contain the neck region corresponding to the 30 minutes dynamic acquisitions. These were followed by the remaining two bed positions with two minutes acquisition time. The resulting whole-body PET/CT scan gives the same static image used in routine treatment planning.

Still during the second imaging session, after 95 minutes of radiotracer injection, another static PET scan was performed only of the neck region with four minutes acquisition time. The two static images of the neck region are appended to the dynamic acquisitions to create a total of 22 dynamic frames.

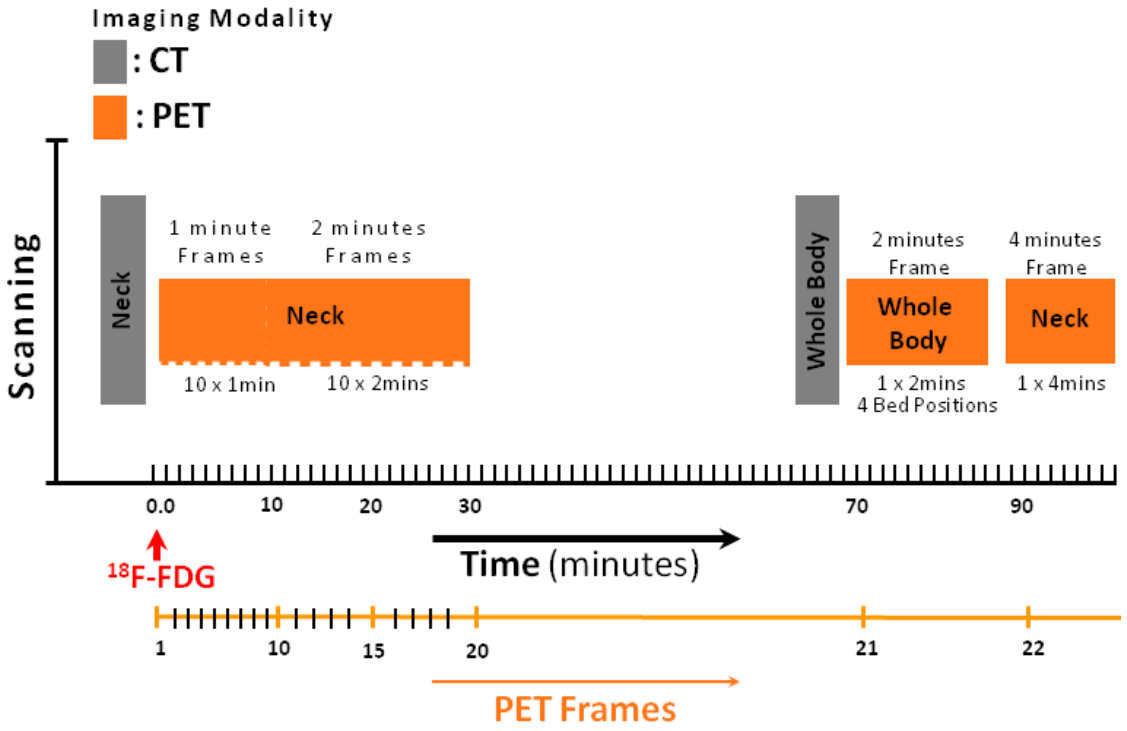


Figure 4: *Diagrammatic representation of the scanning protocol which was designed for this study. An initial CT scan was made for planning purposes of the PET scans and afterwards PET scanning was performed immediately after the injection of radiopharmaceutical drug FDG; twenty PET frames were acquired within the first 30 minutes. Later, in the second scanning session a whole body CT scan was acquired which was part of routine diagnostic scan followed by whole body PET scan. Lastly, a PET acquisition of the neck region was also obtained which made in total twenty two dynamic PET frames of the neck region.*

Image reconstruction was done with an iterative VUE Point Fx method (GE Healthcare, 2011) with 2 iterations and 24 subsets. A reconstruction matrix of 192 x 192 and FOV 70 cm was used. After reconstruction and decay correction, images of the static

acquisition at the 21st time point (70-75 minutes & 4 minutes frame length) were used in order to generate SUV images according to equation 1 (section 2.1.3). We calculated SUV for the entire image used in treatment planning as well as separately for all those voxels which were present within our VOIs.

4.3 Extracting Volumes of Interest

4.3.1 Contouring tumour and inflammation

As mentioned in section 2.5.3, our experiments require independent approximation of tissue classes in an image, for use in classifier evaluation. As a part of patient's normal treatment planning protocol, a qualified radiation oncologist drew 3 dimensional (3D) contour sets around the tumours. These contours were drawn manually, on SUV images from the 21st time frame, as part of clinical treatment planning procedure required for intensity modulated radiation therapy (IMRT) and also used as the best available 'truth' in classifier evaluation.

All of the patients underwent dental extractions before the start of their treatment because periodontal infections during cancer therapy may lead to significant morbidity and mortality (Epstein and Moore, 2001); therefore, patients also had ^{18}F-FDG uptake at the site of dental extraction due to the presence of inflammation at those sites and these areas could also be used as inflammatory VOIs. The radiation oncologist also drew contours around these dental inflammatory regions of uptake.

All of these contours were drawn on the spatially co-registered PET/CT images because CT has higher resolution and provides better anatomical visualisation and demarcation than PET imaging (Gordin et al., 2007); the corresponding PET voxels of these CT contours were identified (Figure 5 A/B) and subsequently used in our research.

4.3.2 Automatic method of healthy tissue definition

As part of this thesis research, a method was devised for isolating PET voxels surrounding the tumours, providing 'ground truth' for the class of normal tissue immediately outside of the tumour. This is an important class because, for the purpose of automated GTV contouring, it is exactly these voxels that should be distinguished from tumour. These voxels were isolated by automatically defining a three dimensional regular shaped bounding box around the irregular shaped tumour, to enclose all tumour voxels, and subtracting those voxels from bounding box which belonged to tumour class (Figure 5 C). Finally, those voxels were also excluded, which had high tracer uptake and were included in other nearby VOI contours (e.g. inflammation, lymph node) thus representing any other metabolic process.

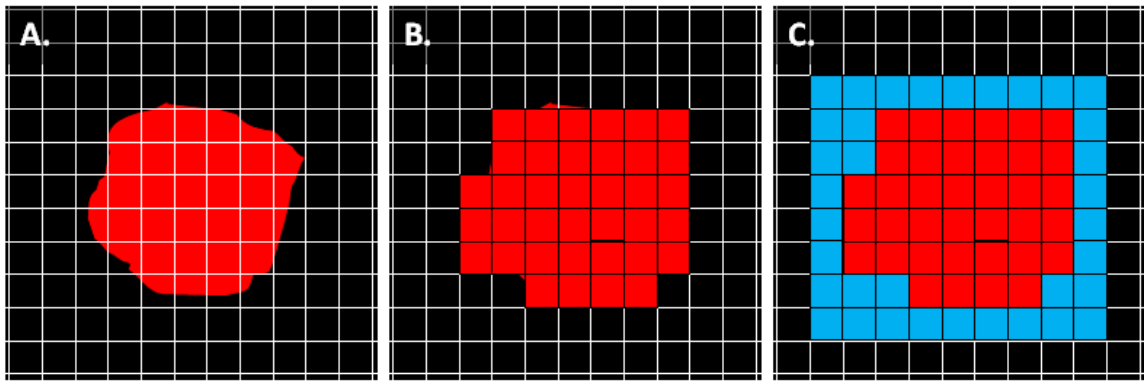


Figure 5: *Procedure (shown only in 2 dimensions) of extracting voxels from volume of interest contours. A. All initial contours were delineated on the CT scan images B. Process of the identification of the corresponding PET imaging voxels as that of particular VOI contours. C. The automatic method of acquiring the PET voxels immediately surrounding the tumours (red) representing normal tissue (blue).*

As well as healthy tissue around tumours, physiological uptake that was observed at the site of palatine tonsil in the case of one patient was also delineated using an automated method; a regular shaped three dimensional bounding box was defined manually around the area of physiological uptake and afterwards those voxels were excluded which had less than 50% of the maximum SUV (SUV_{max}).

For the five head and neck patients included in our study, 5 tumour VOIs, 8 inflammation VOIs, 5 healthy tissue VOIs and 1 physiological uptake (palatine tonsil) VOI were defined both by manual as well as automated delineation methods. These delineations provide 'ground truth' for training and evaluating classification experiments, as well as validation by correlation with clinical data. However, it should also be noted that manual contours do have inevitable expert subjectivity and tendency to include some ambiguous tissue near the boundary in order to avoid under-segmentation. All these VOIs are mentioned in table 2 along with their total volume in cm^3 and number of voxels in PET space. Moreover, the anatomical locations of these VOI have been shown in figure 5 and Appendix A includes all the fused PET-CT images of these VOIs along the craniocaudal axis. All inflammation classes are labelled as left- or right-inflammation according to location in the image (e.g. 'left inflammation' in figure 6 A ii) rather than the patient's frame of reference.

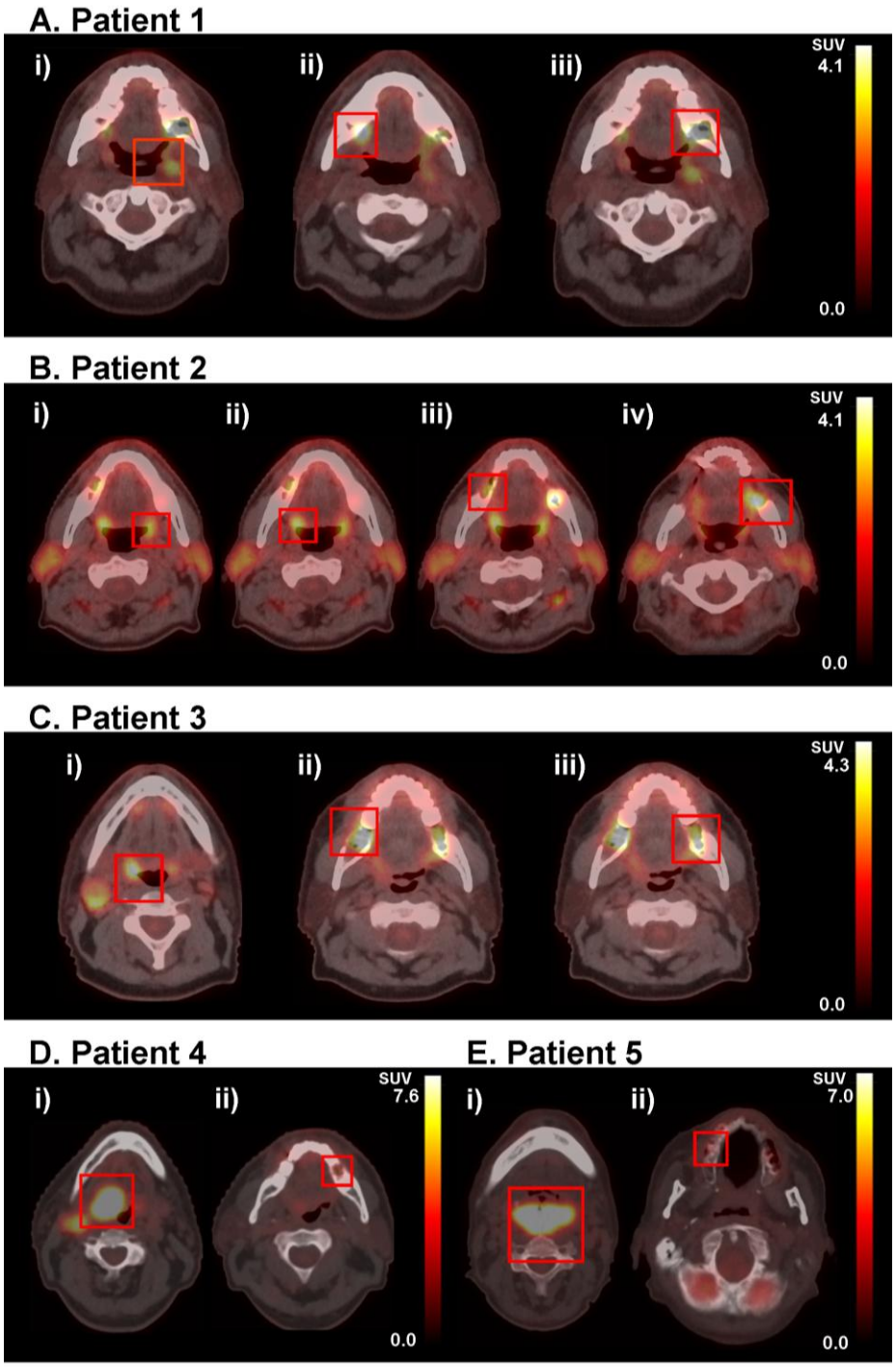


Figure 6: *Fused PET-CT images of the patients showing anatomical location of the VOIs included in the study. A. Patient 1 images, red coloured bounding box encloses i) tumour located at palatine tonsil, ii) inflammation within right jaw at the site of dental extraction, iii) inflammation within left jaw at the site of dental extraction. B. Patient 2 images, red coloured bounding box encloses i) tumour located at palatine tonsil, ii)*

normal physiological uptake of radiotracer within palatine tonsil, iii) inflammation within right jaw at the site of dental extraction iv) inflammation within left jaw at the site of dental extraction. C. Patient 3 images, red coloured bounding box encloses i) tumour located at palatine tonsil, ii) inflammation within right jaw at the site of dental extraction iii) inflammation within left jaw at the site of dental extraction. D. Patient 4 images, red coloured bounding box encloses i) tumour located at base of tongue, ii) inflammation within left jaw at the site of dental extraction. E. Patient 5 images, red coloured bounding box encloses i) oropharyngeal carcinoma, ii) inflammation within left right at the site of dental extraction.

Table 2: *Details of all VOIs included in the study acquired by manual as well as automatic contouring techniques*

Patient Number	VOI	Number of Voxels	Total Volume in cm^3
1	Tumour	201	9.97
	Inflammation Left	33	1.79
	Inflammation Right	36	1.64
	Healthy Tissue	255	12.55
2	Tumour	106	5.26
	Inflammation Left	9	0.45
	Inflammation Right	39	1.93
	Healthy Tissue	134	6.59
	Physiological Uptake	96	4.72
3	Tumour	237	11.76
	Inflammation Left	32	1.58
	Inflammation Right	26	1.29
	Healthy Tissue	225	11.07
4	Tumour	524	25.99
	Inflammation Right	12	0.60
	Healthy Tissue	569	28.01
5	Tumour	1488	73.80
	Inflammation Left	10	0.50
	Healthy Tissue	1742	85.74

4.4 Dynamic Features Computation from TACs

From the generated TACs (one per voxel) a number of proposed features were calculated with the aim of differentiating TACs belonging to different tissue groups based on their shape, slope or other dynamic properties. The properties that were chosen for investigation are those that seemed to differ between classes in training data by visual inspection, or that may reasonably be expected to be influenced by different mechanisms of glucose transport. In total we developed seven dynamic features denoted D1, D2, D3...D7. Each of these feature's mathematical derivation is described below, along with possible relevance to the biological characteristics of tissues following on from knowledge of the mechanisms described in section 2.3. All of these computations were performed in the programming software Matlab (Version: R2006b).

4.4.2 D1: Retention Index

Retention Index has been used earlier in various dual time point ^{18}F-FDG studies which have been described in section 2.1.3. It represents fractional or percentage change in the activity concentration between an early time point and late time point of the PET scan. Findings of Higashi et al., (2002) suggested that the retention index can be an indicator of hexokinase II expression and it might be an indicator of phosphorylation rate in the area of interest. In our experiments we calculated the retention index using the following formula,

$$Retention\ Index\ = \frac{A_{late} - A_{early}}{A_{early}}$$

(Equation 3)

Where, A is the activity concentration at a specific point of time. As mentioned in section 4.2, our scanning protocol had 22 PET frames; for A_{late} we used the activity concentration in the 21st PET frame and for A_{early} we used the mean activity concentration of the first 10 PET frames.

4.4.2 D2: Early Slope

The slope of the TAC in the first 30 minutes was also used as a feature which may have the possibility to differentiate between tissues of different predominant function but similar long term uptake. Early slope of the TAC was calculated using following the formula of simple slope,

$$Early\ Slope\ = \frac{mean(A_{15}, A_{16} \ldots A_{20}) - mean(A_1, A_2 \ldots . A_{10})}{mean(T_{15}, T_{16} \ldots T_{15}) - \ mean(T_1, T_2 \ldots T_{10})}$$

(Equation 4)

where T is the time of that PET frame.

4.4.3 D3: Area under the $TAC_{20\text{-}21}$

The third feature calculated was the area under the time activity curve between the 20th PET frame (30th minute) and the 21st PET frame (~70th minute). Area under the curve between these two time points was computed using the following formula, and graphical explanation of the equation has been shown in figure 7.

$$Area\ under\ the\ TAC_{20-21}\ = \frac{(T_{21} - T_{20}) \times (A_{21} - A_{20})}{2} + B$$

(Equation 5)

where B is equal to $(T_{21} - T_{20}) \times A_{20}$ (see figure 7 for graphical explanation)

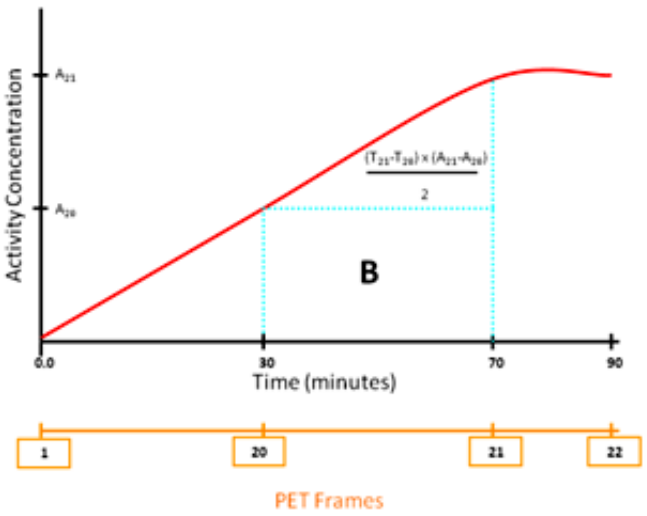


Figure 7: *Graphical representation of the mathematical derivation of D3 feature where area under the TAC of 20 and 21 PET frame is computed. Area under the* $TAC_{20\text{-}21}$ *is the sum of* $[(T_{21}\text{-}T_{20}) \times (A_{21}\text{-}A_{20})/2]$ *and B; where B is equal to* $[(T_{21}\text{-}T_{20}) \times A_{20}]$.

4.4.4 D4: Sum Fluctuation$_{1\text{-}22}$

Fluctuations in the activity concentration in TAC at a particular voxel might be representing flux in and out of the ^{18}F-FDG radiotracer within that particular corresponding area; and the sum of all these fluctuations could possibly represent total radiopharmaceutical flux in-out capacity of the corresponding tissue. Where fluctuations (changes over small time differences) are small, it might be argued that they are largely governed by imaging noise; however, it should be noted that different VOIs belonging to the same PET image will suffer equal amounts of imaging noise and therefore features based on summing all these fluctuations is technically justified. The sum of fluctuations was computed using the following formula,

$$Sum\ Fluctuation_{1-22} = |A_2 - A_1| + |A_3 - A_2| + |A_4 - A_3| \ldots\ldots + |A_{22} - A_{21}|$$

(Equation 6)

4.4.5 D5: Variance of local change

Next we quantify the variance in the local change in activity concentration, over the extended period of time. This variance within a single PET voxel, is one way to characterise the overall shape of the TAC from that voxel, governed by the in-flux and out-flux of the radiotracer within the corresponding tissue. This feature is unique in the sense that it does not depend upon the final accumulative activity concentration within a tissue; rather, it represents the capacity of the corresponding tissue for the in-flux and out-flux of the radiotracer. For the purpose of calculation is variance Matlab's built-in function 'var' was used.

$$Variance\ of\ local\ change = Var[(A_2 - A_1), (A_3 - A_2), \dots (A_{22} - A_{21})]$$

(Equation 7)

where A_n is the activity concentration within the voxels at the n^{th} PET frame.

4.4.6 D6: Temporal Variance

Temporal variance of the entire time activity curve represents the total change in the radioactivity concentration over the entire period of time. The greater the temporal variance, the greater the activity concentration gained over the period of the whole TAC. This variance also reflects how quickly the signal becomes stable. However, it is one limitation of this feature that it cannot differentiate between those TACs which show early uptake and those which show greater activity concentration during the later time period. Temporal variance within a single TAC of each voxel was computed using Matlab's built-in function for the calculation of the variance within a given sample,

$$Temporal\ Variance = Var[A_1, A_2, A_3 \dots A_{22}]$$

(Equation 8)

where A_n is the activity concentration within the voxels at the n^{th} PET frame.

4.4.7 D7: Sum of three slopes

As earlier mentioned, the slope of a TAC represents the trapping of ^{18}F-FDG within the particular tissue. Three slopes of the entire TAC were individually calculated and afterwards summed. Graphical explanation of the feature has been shown in figure 8.

$$Sum\ of\ three\ slopes = Slope\ 1 + Slope\ 2 + Slope\ 3$$

(Equation 9)

where,

$$Slope\ 1 = \frac{mean(A_{15}, A_{16} \ldots A_{20}) - max(A_1, A_2 \ldots . A_5)}{mean(T_{15}, T_{16} \ldots T_{15}) - max(T_A)}$$

(Equation 9.1)

$$Slope\ 2 = = \frac{A_{21} - A_{20}}{T_{21} - T_{20}}$$

(Equation 9.2)

$$Slope\ 3 = = \frac{A_{22} - A_{21}}{T_{22} - T_{21}}$$

(Equation 9.3)

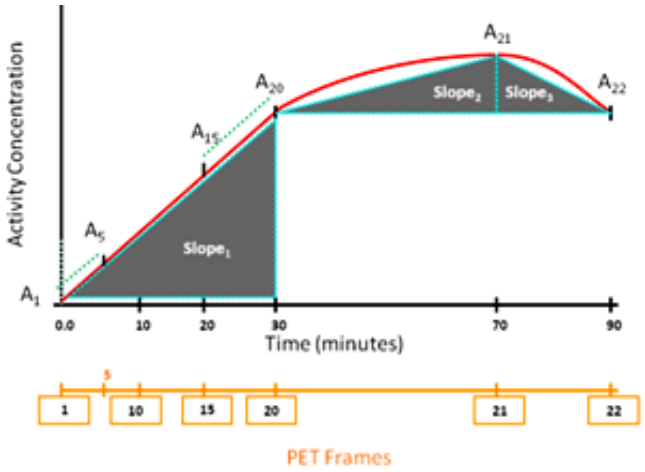


Figure 8: *Graphical representation of the mathematical derivation of D7 feature where sum of three slopes over the TAC is computed. Slope 1 is the slope of the TAC over first 20 frames; slope 2 is the slope between 20th and 21st PET frame and slope 3 is the slope between 21st and 22nd PET frame*

4.5 Feature Classification and Classifier Performance

To classify the features calculated for each of the voxels included in our study, we used GMM and K-means classification algorithms. These algorithms have already been described in section 2.5. First we applied the classification algorithm to a mixture of data belonging to two different classes, from the set of four namely tumour, healthy, inflammation and physiological uptake. The algorithms were set up to divide the mixture into two classes, i.e. the number of classes is known but the truth labels are not. Next, the algorithms assigned data with a label according to whether it belongs to the positive class (e.g. tumour) or negative class (e.g. healthy tissue).

For evaluating the performance of the GMM algorithm, the labels produced by GMM were compared with truth labels using ROC analysis (section 2.5). The controlling parameter, which was varied in order to create the multiple points of a ROC curve, was simply a threshold on the feature value. Finally, area under the ROC curve (AUC_{ROC}) was considered as a measure for GMM classifier performance. AUC_{ROC} of maximum 1.0 can only be achieved if there is perfect match of truth labels with all labels produced by algorithm.

In the case of the K-means algorithm, the truth labels were compared with the label produced by the algorithm and afterwards sensitivity and specificity were calculated using the following formulae,

$$Sensitivity = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives\ + Number\ of\ false\ negatives}$$

(Equation 10)

$$Specificity = \frac{Number\ of\ true\ negative}{Number\ of\ true\ negatives + Number\ of\ false\ positives}$$

(Equation 11)

All the classification and performance evaluation were also implemented in the programming software Matlab (Version: R2006b). K-means classification is a built-in

functions of the Matlab statistical toolbox package; whereas, GMM code was adapted from code taken from Simulink's on-line archive of open-source libraries (Matlab Central – File exchange, 2013).

Classifier performance of different computed dynamic features were also compared with the classification performance of the static parameter SUV. For each binary classification (e.g. tumour vs. normal tissue in the same patient) the performance (AUC_{ROC} value in the case of GMM or sensitivity/specificity in the case of K-means) of feature and SUV-based classification was compared using Wilcoxon rank-sum tests. These tests investigate the difference between classifier performances over the whole group of patients, and are preferred over a parametric test (e.g. Student's t-test) for cases where the number of patients is low. We used the Wilcoxon test for those features which were showing apparent superior discriminatory power as compared to SUV, in order to reveal any statistically significant differences.

4.6 Correlation with Complementary Data

It was interesting to see if a feature derived from the dynamic data correlated with the information that has been obtained independently of the dynamic PET imaging protocol. We also tried to discover such correlations between the dynamic features and clinical complimentary data which includes tumour grade from histopathology, HPV status and T-stage (from TNM staging, discussed in section 2.2) of the patient. Such correlations provide extra evidence of the affinity of the image analysis methods for measuring molecular properties and any correlation could in principle mean that the expensive, invasive procedure of biopsy and genetic analysis could be replaced by further analysis of the data available from PET studies.

Our complimentary data were either binary (HPV status) or ordinal (stage, grade) and was available for only five patients; therefore, we decided to use Wilcoxon rank-sum test to seek any significant difference between mean dynamic features of the entire tumour VOI belonging to different sub-categories (HPV, grade, stage).

5. Experiments and Results

5.1 Data preparation

Each of the voxels included in a VOI had a series of PET acquisition's activity information along the set time profile which could be used to plot time activity curves (TACs); therefore, we plotted these activity intensities, obtained from PET images, along the time profile in order to see different trends of radiotracer dynamics in metabolically different tissues. These different time activity curves are shown in Figure 9. Time activity curves provided us with information on the dynamics of tracer uptake and washout in the area of interest. As mentioned earlier, we assumed that the time activity curves of the area of inflammation, tumour and healthy tissue will be different because of different rate of metabolic activity & different metabolic pathways.

In the TACs it is interesting to observe that the tumour class of the first three patients having tonsilar carcinoma has generally less activity concentration compared to the inflammatory class; while in the last two patients (non-tonsilar carcinomas) the tumours are showing very high activity concentration, higher than inflammatory regions. Moreover, visual inspection of TACs reveals that the shape of the time activity curve of the tumour class in the first three patients is different from the last two.

The SUV for each of the voxels was calculated using the 21st PET frame (70-75 minutes after radiotracer injection) using equation 1 (section 2.1.3). Mean SUV for the VOI was calculated by averaging all the SUVs present within a volume of interest, and similarly for the mean retention index. Table 3 shows the mean SUV and mean retention index for all of the VOIs included in our study. It is interesting to note that the tumour VOI of patients 4 and 5 (non-tonsilar carcinomas) have remarkably higher SUVs compared to the other three patients having tonsilar carcinomas.

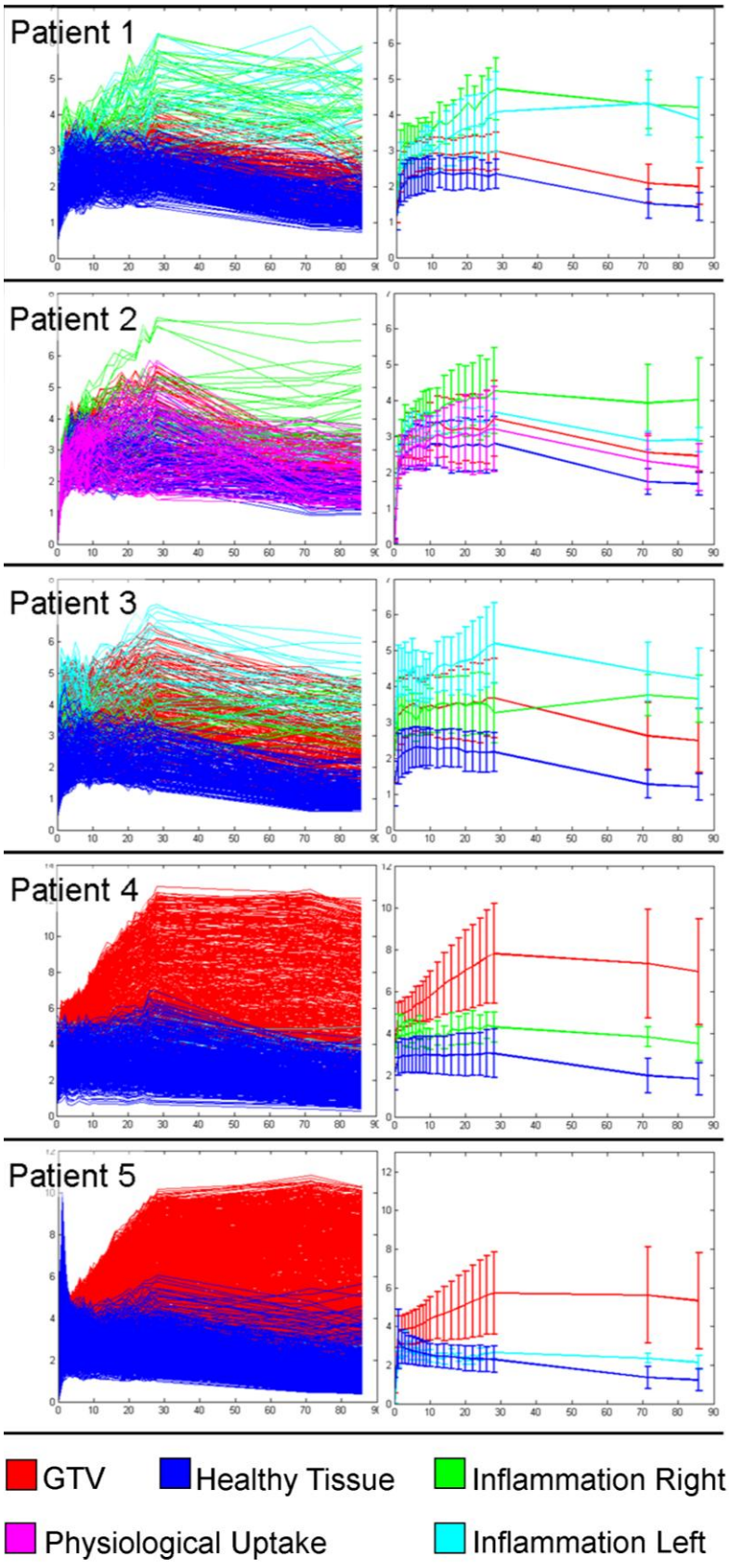

Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
GTV
Healthy Tissue
Inflammation Right
Physiological Uptake
Inflammation Left

Figure 9: *Time Activity Curves were plotted for each of the voxels included in our study; the graphs on the left side show the raw TACs of all the voxels included in different VOIs, different colours represent their origin from a particular class, with colour coding explained in the key. The graphs on the right show the same averaged TAC belonging to different VOI along with error bars given by ± 1 standard deviation.*

Table 3: *Mean SUV and mean retention index of all the VOI included in the study.*

Patient Number	VOI	SUV_{mean}	$R.I_{mean}$
1	Tumour	2.08	-23.24
	Inflammation Left	4.33	53.16
	Inflammation Right	4.29	31.14
	Healthy Tissue	1.51	-31.67
2	Tumour	2.56	-10.53
	Inflammation Left	3.94	-5.19
	Inflammation Right	2.88	19.45
	Healthy Tissue	1.73	-32.17
	Physiological Uptake	2.30	-11.14
3	Tumour	2.62	-22.89
	Inflammation Left	3.76	2.01
	Inflammation Right	4.41	13.57
	Healthy Tissue	1.28	-42.62
4	Tumour	7.32	40.44
	Inflammation Right	3.82	-5.68
	Healthy Tissue	1.98	-32.29
5	Tumour	5.61	41.40
	Inflammation Left	2.35	-10.79
	Healthy Tissue	1.36	-51.59

After the data preparation, a series of different experiments were performed on the available data; all the experiments were divided into three major groups, based on the VOIs to be compared,

- a. GTV and Inflammation experiments
- b. GTV and normal tissue surrounding GTV experiments
- c. GTV and Normal Physiological Uptake experiments

The following subsections describe each of the performed experiments as well as their results; implications of the obtained results will be further analysed in the discussion section.

5.2 Discriminating GTV from Inflammation

Experiments for discriminating GTV from inflammation were set up to address hypothesis $\mathcal{H}1$ (section 3.2).

Gaussian Mixture Model classification was performed and area under the ROC curve was used as a combined measure for the sensitivity and specificity of the particular feature in successfully differentiating between two classes (Figure 11).

Next, a K-means clustering algorithm was also utilised to automatically distinguish two classes. Performance of the K-means algorithm was evaluated in terms of sensitivity and specificity.

We performed these experiments for all pairs of GTV and inflammation, and used a Wilcoxon rank-sum test to look for significant differences in performance between dynamic features and SUV, considering a significant cut off P-value of 0.05.

Figure 10 shows the flowchart of all the experiments performed for differentiating GTV and inflammatory volumes, where D*i* is the dynamic feature calculated over the VOI and the same data flow is repeated for '*i*' from 1 to 7 (D1, D2, D3...D7). Figure 11 graphically represents the results of the GMM algorithm and Figure 12 shows the results of the K-means algorithm, each showing selected results of the Wilcoxon test.

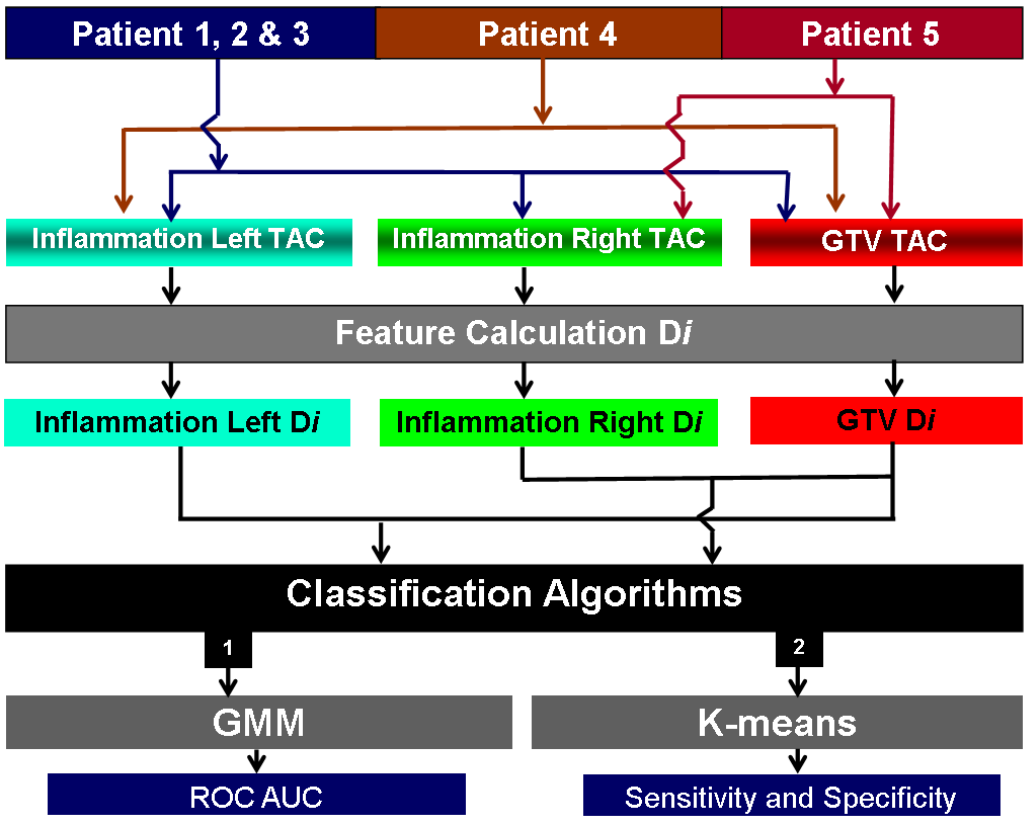
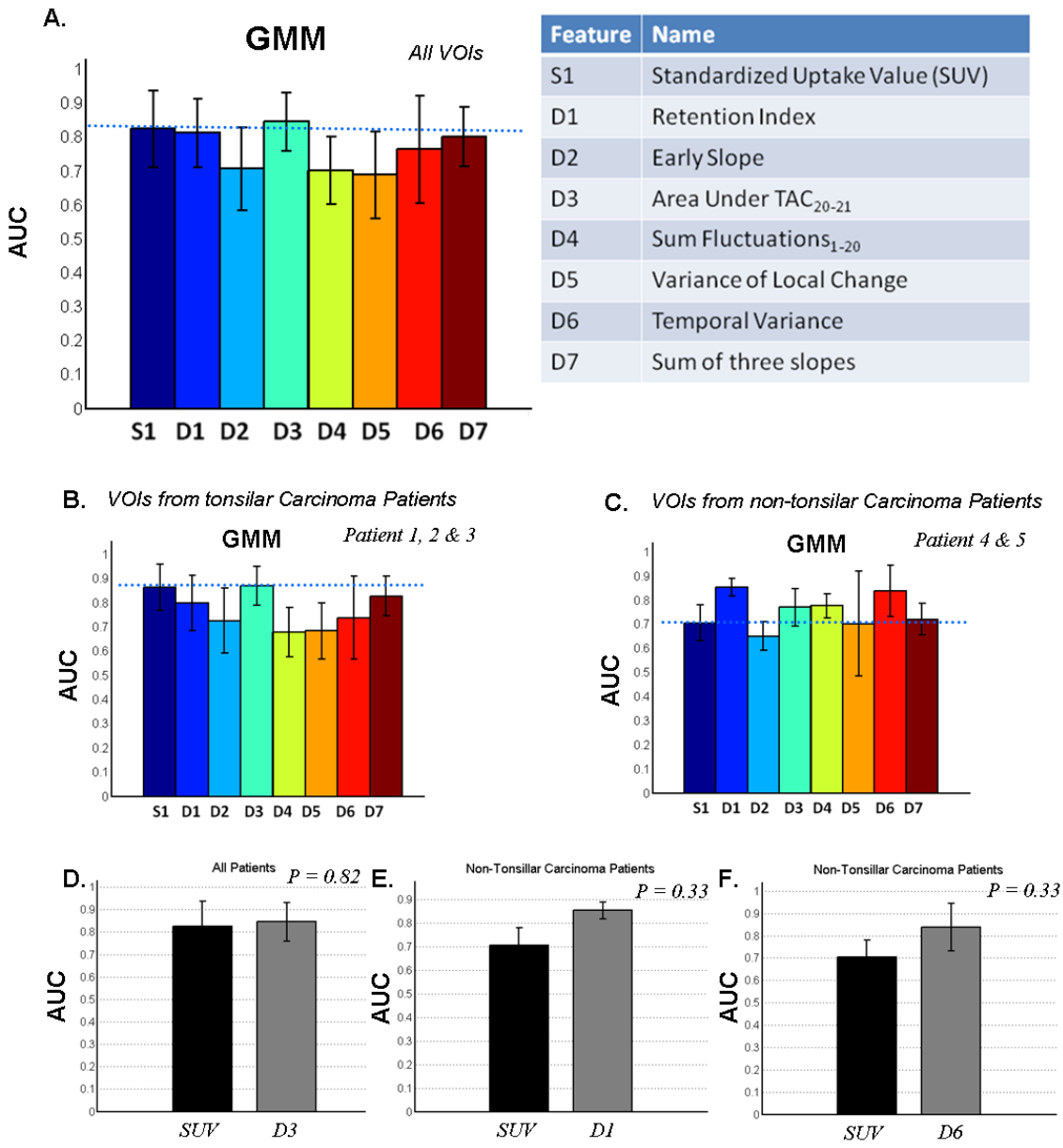


Figure 10: *Flowchart of the processes involved in the experiments of discriminating GTV from Inflammation.*



Feature	Name
S1	Standardized Uptake Value (SUV)
D1	Retention Index
D2	Early Slope
D3	Area Under TAC_{20-21}
D4	Sum $Fluctuations_{1-20}$
D5	Variance of Local Change
D6	Temporal Variance
D7	Sum of three slopes

Figure 11: *GMM Algorithm results of discriminating between GTV and inflammation A: Showing aggregated results, along with error bars, in the form of ROC area under curve in all patients VOIs (In total 8 comparisons using one feature). The first navy blue coloured bar shows the ability of static parameter S1 (SUV) to successfully classify two classes while other bars show the ability of the dynamic features D1-7. B: Showing the performance of GMM classification algorithm in the patients having palatine tonsil carcinoma. C: Showing the performance of GMM classification algorithm in the patients having non-tonsil carcinoma (base of the tongue and oropharyngeal*

carcinoma). The horizontal dotted blue line in A, B & C indicate level of static parameter S1 (SUV) to successfully classify two classes. D: Wilcoxon rank-sum test between the obtained AUC values of SUV and dynamic feature D3 in all of the patient's VOIs (P=0.82). E: Wilcoxon rank-sum test between the obtained AUC values of SUV and dynamic feature D1 in VOIs of non-tonsil carcinoma patients (P=0.33). F: Wilcoxon rank-sum test between the obtained AUC values of SUV and dynamic feature D6 in VOIs of non-tonsil carcinoma patients (P=0.33). Error bars in all of the graphs represent ± 1 standard deviation.

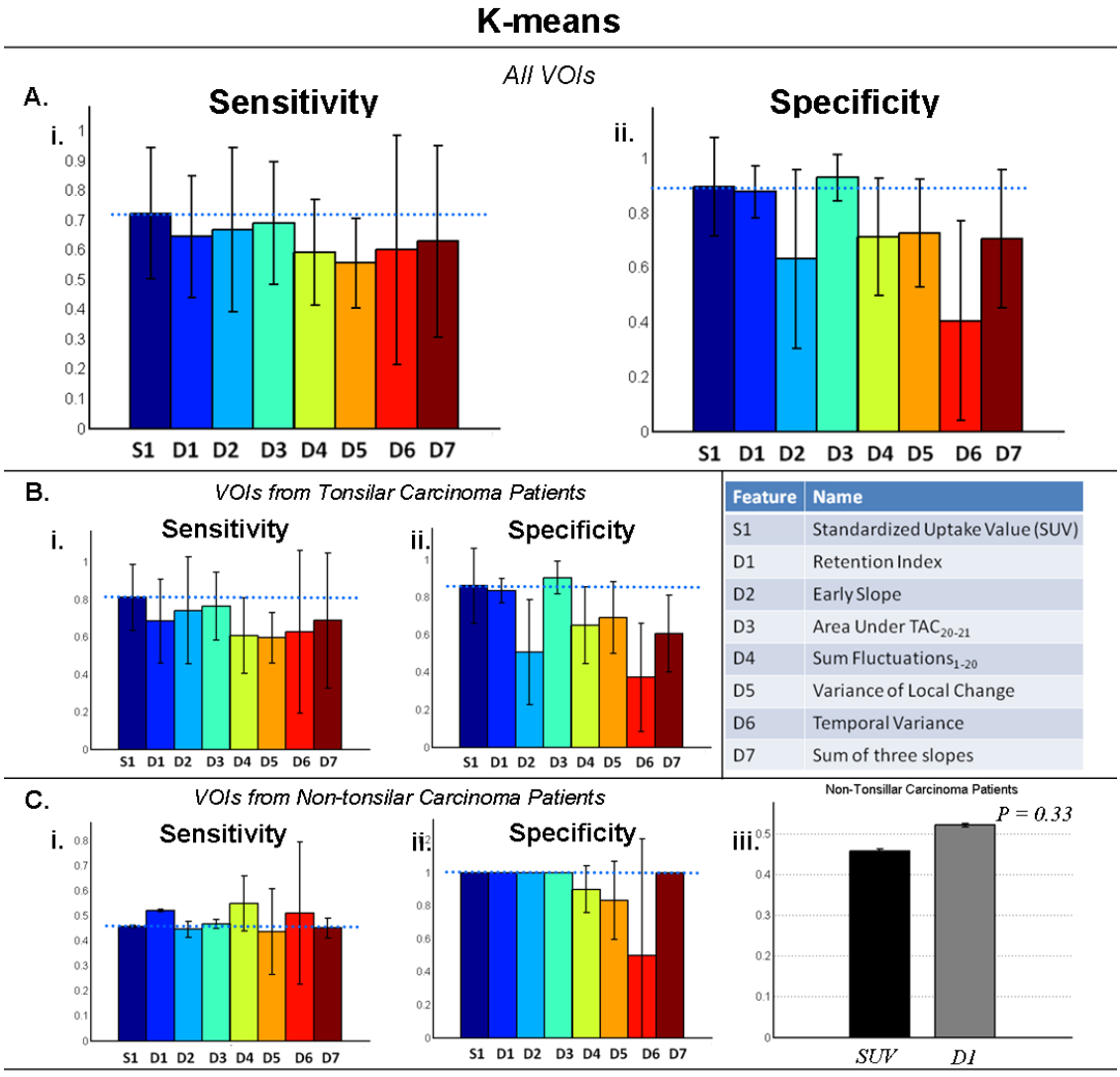


Figure 12: *K-means algorithm results of discriminating between GTV and inflammation A: Showing aggregated results, along with error bars, in the form of sensitivity (i) and*

specificity (ii) in all patients VOIs (in total 8 comparisons using one feature) and analysing 7 dynamic features. The first navy blue coloured bar shows the ability of static parameter S1 (SUV) to successfully classify two classes while other bars show the ability of the dynamic features. B: Performance of K-means classification algorithm in the patients having palatine tonsil carcinoma. C: Performance of K-means classification algorithm in the patients having non-tonsil carcinoma (base of the tongue and oropharyngeal carcinoma). (i) sensitivity of the algorithm in this subgroup (ii) specificity of the algorithm in this subgroup (iii) Wilcoxon rank-sum test between the sensitivities of SUV and dynamic feature D1 in patient's VOIs of non tonsil carcinoma (P=0.33). The horizontal dotted blue line in the graphs represent level of static parameter S1 (SUV) to successfully classify two classes and error bars indicate ± 1 standard deviation.

The lower values of the area under the ROC curve (AUC) and lower values of sensitivity & specificity indicates that the classifier struggles to discriminate between inflammation and the GTV. It is interesting to observe that some developed features have better performance using one classification algorithm than the other one. Dynamic feature 3 (D3) has shown better discriminatory ability using GMM as compared to K-means algorithm in patients with non-tonsil carcinoma.

As mentioned in section 4, manual contours drawn by a radiation oncologist introduce inevitable expert subjectivity and there may be tendency to include some ambiguous tissue near the boundary in order to avoid under-segmentation; therefore VOIs from manual contours belonging either to GTV or inflammatory region may also contain some normal tissue voxels. As a result, ROC or sensitivity/specificity analysis using manual ground truth can give biased evaluation of classifier performance. In order to minimise the effect of subjective ground-truth we repeated classifier evaluation using ground truth obtained from an SUV thresholding technique. Starting with the manual contours (both GTV and inflammation), we thresholded those voxels which were showing less than 60% of the SUV_{max}; the voxels above this threshold were considered

to be purely representing one metabolic process (either cancer or inflammation). Afterwards, we used both of our classification algorithms to compare these two classes (GTV and inflammation). Results of these experiments are shown in figure 13.

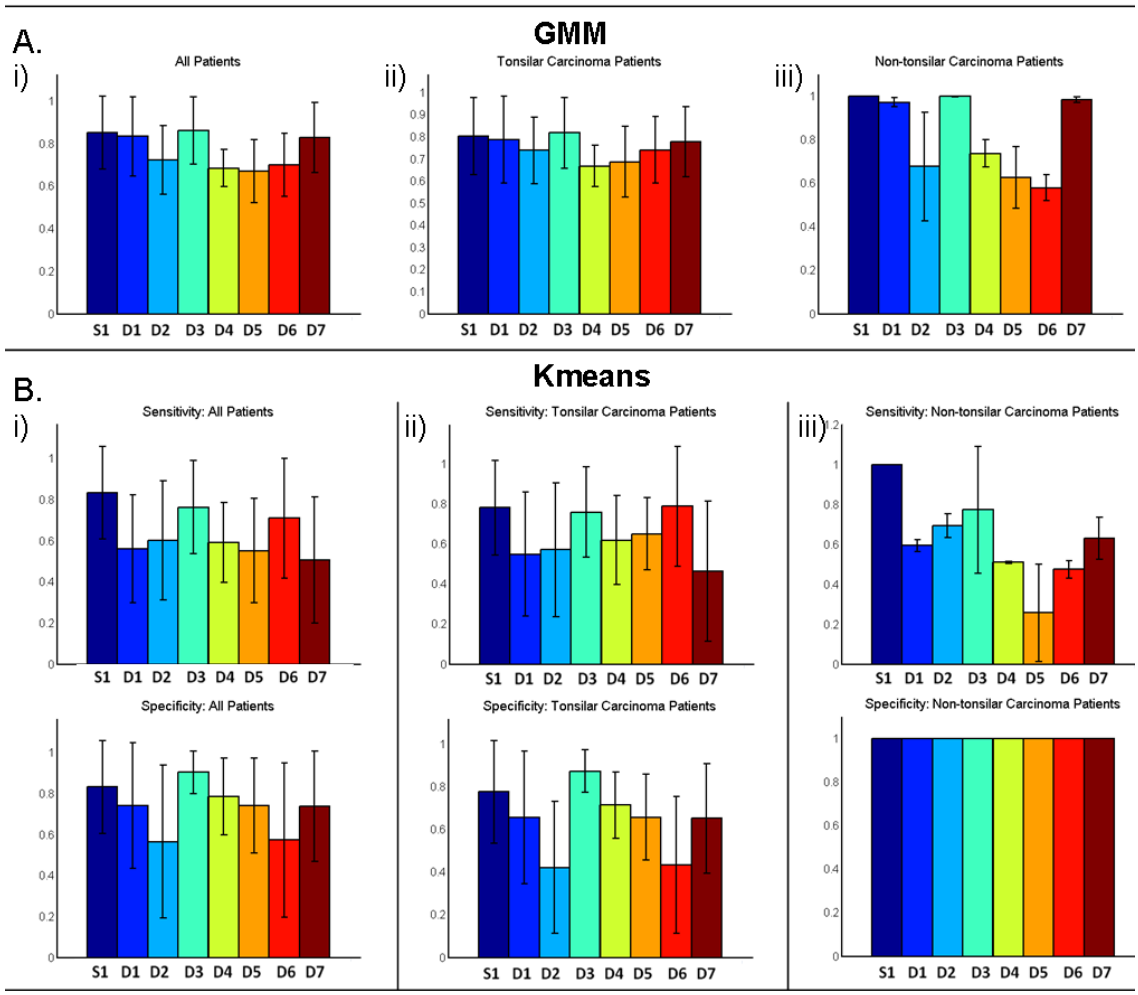


Figure 13: *Results of both GMM algorithm and K-means algorithm in discriminating GTV and inflammation after thresholding those voxels from manual contours which were showing less than 60% of* SUV_{max}*. A: GMM algorithm results in terms of AUC obtained from ROC analysis (i) results of all patients' VOIs (ii) results of comparison of VOIs belonging to tonsil carcinoma patients (iii) results of comparison of VOIs belonging to non-tonsil carcinoma patients. B: K-means algorithm results in terms of sensitivity and specificity (i) results of all the patients' VOIs (ii) results of comparison of VOIs belonging to tonsil carcinoma patients (iii) results of comparison of VOIs belonging to non-tonsil carcinoma patients.*

It is interesting to observe in the results of the experiments performed using SUV threshold-based ground truth that the overall discriminatory power of almost all of the features increased; however, the overall relative trend in between the features remained the same.

In all of these experiments some of the dynamic features showed better performance as compared to static parameter SUV (figure 11D, E, F & figure 12C iii); however, they were not able to provide any significantly superior performance considering a confidence level of 95%. Therefore, despite we found evidence in favour of the hypothesis, hypothesis $\mathcal{H}1$ cannot be formally accepted.

5.3. Discriminating GTV from Surrounding Tissue

Clustering experiments were also performed for discriminating the GTV class from voxels immediately surrounding the tumours. Hypothesis $\mathcal{H}2$, mentioned in section 4, led all these experiments. The same dynamic features were utilised to assess their usefulness in discriminating two classes using GMM & K-means clustering algorithms. Figure 14 shows a flowchart of the steps taken for these experiments, where D*i* is the feature calculated over the VOI and the same data flow is repeated for '*i*' from 1 to 7 (D1, D2, D3...D7); figure 15 represents all of the results obtained.

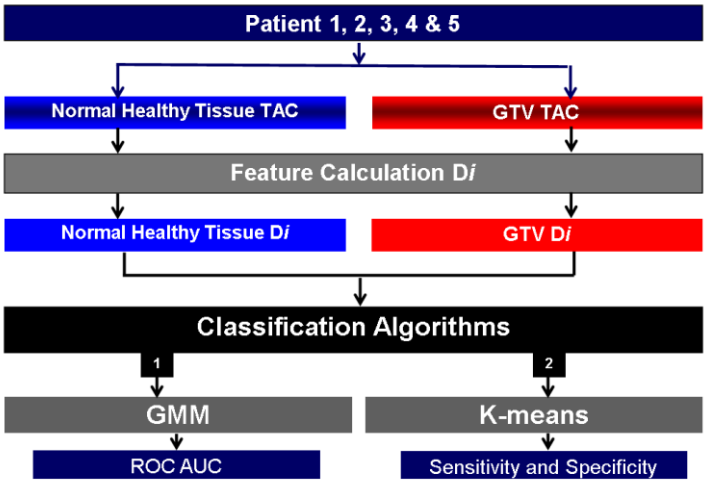
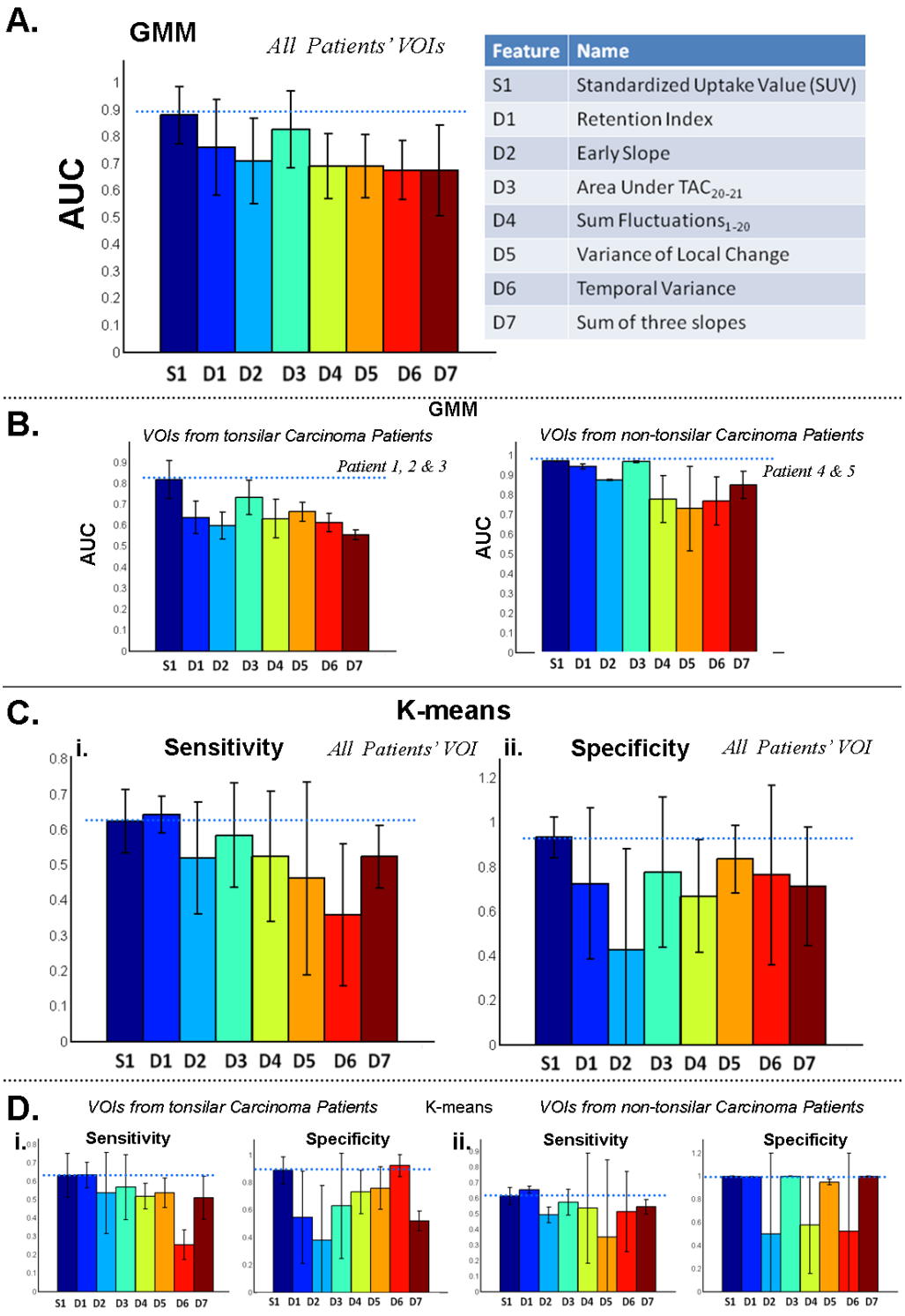


Figure 14: *Flowchart of the processes involved in the experiments of discriminating GTV from the healthy tissue surrounding tumour.*



Feature	Name
S1	Standardized Uptake Value (SUV)
D1	Retention Index
D2	Early Slope
D3	Area Under TAC_{20-21}
D4	Sum $Fluctuations_{1-20}$
D5	Variance of Local Change
D6	Temporal Variance
D7	Sum of three slopes

Figure 15: *A: Results, along with error bars indicating ± 1 standard deviation, in the form of ROC area under curve of the Gaussian Mixture Model clustering in all of patients. The first navy blue coloured bar shows the ability of static parameter S1*

(SUV) to successfully classify two classes while other bars show the ability of the dynamic features. B: Performance of GMM classification algorithm in sub-groups, graph on left: patients having palatine tonsil carcinoma; graph on right: patients having non-tonsilar carcinoma of head and neck region. C. Performance of K-means classification algorithm in terms of sensitivity and specificity of the developed features to discriminate between two classes (GTV and Healthy tissue surrounding tumour)in all patients' VOIs. D: Performance of K-means classification algorithm in sub-groups (i) patients having palatine tonsil carcinoma; (ii) patients having non-tonsilar carcinoma of head and neck region. The horizontal dotted blue line in all of the graphs indicates that level of static parameter S1 (SUV) to successfully classify two classes.

As the tumour contours have been manually drawn in CT space while keeping in mind PET uptakes on the SUV scale, it is therefore quite understandable that SUVs are showing better discriminatory power in both of the algorithms. However, it should also be noticed that SUVs are not showing maximum AUC or sensitivity/specificity of 1 because SUVs are not the only parameter for tumour manual delineations by radiation oncologists. Possible application for these experiments could have been the automated tumour delineations based on dynamic features mimicking the manual contouring of radiation oncologists; and differentiation of tissue based on metabolic processes was not the prime objective; therefore, experiments with SUV thresholding technique were not performed in this subcategory of experiments. Keeping in view the results of these experiments we can deduce that hypothesis $\mathcal{H}2$ cannot be formally accepted.

5.4 Discriminating GTV from Normal Physiological Uptake

In one of our tonsillar carcinoma patients (patient 2), there was bilateral FDG uptake in the palatine tonsil region; however, biopsy findings revealed the presence of tumour in the left side, while uptake on the contra-lateral side was classified as normal physiological uptake (Figure 16). We hypothesised that this similar looking normal physiological uptake can be differentiated from the tumour's uptake based on our

developed dynamic features (Hypothesis $\mathcal{H}3$). Figure 17 show the flowchart of all the steps taken for these experiment where *Di* is the feature calculated over the VOI and the same data flow is repeated for '*i*' from 1 to 7 (D1, D2, D3...D7); figure 18 show all of the obtained results.

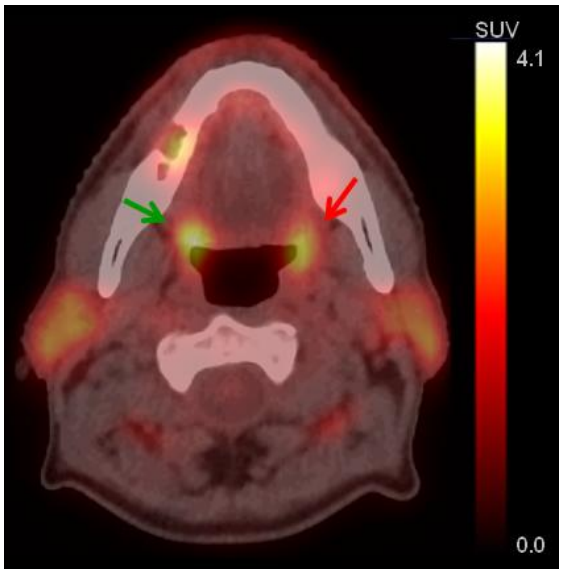


Figure 16: *Hybrid PET-CT static image of one of the patients (patient 2), included in our study showing multiple FDG uptake areas which can be visualised as bright spots on the SUV scale. Red arrow pointing bright spot as result of FDG uptake due to tumour in the palatine tonsil area (confirmed by biopsy findings) while the green arrow showing normal physiological uptake in the contra-lateral tonsil region.*

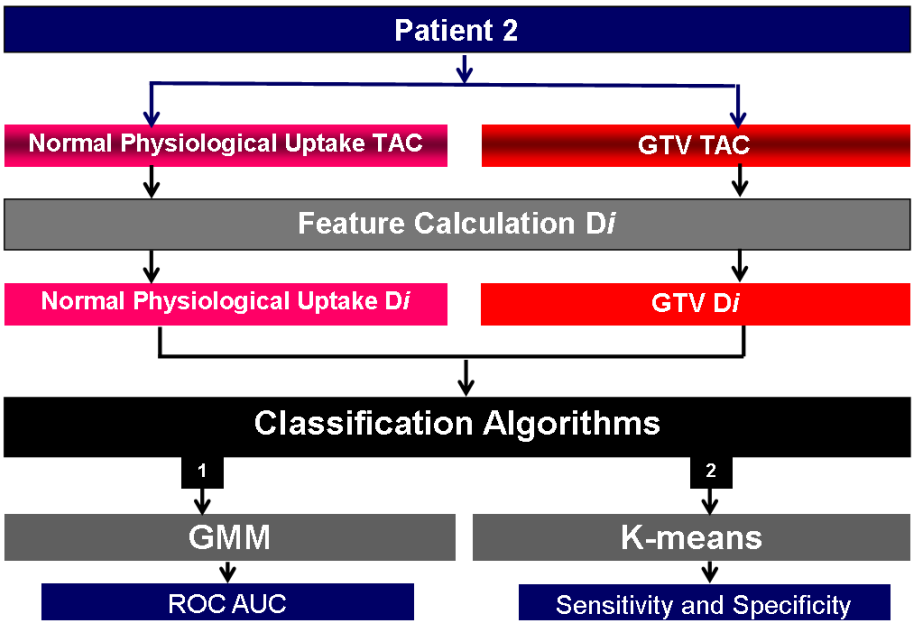
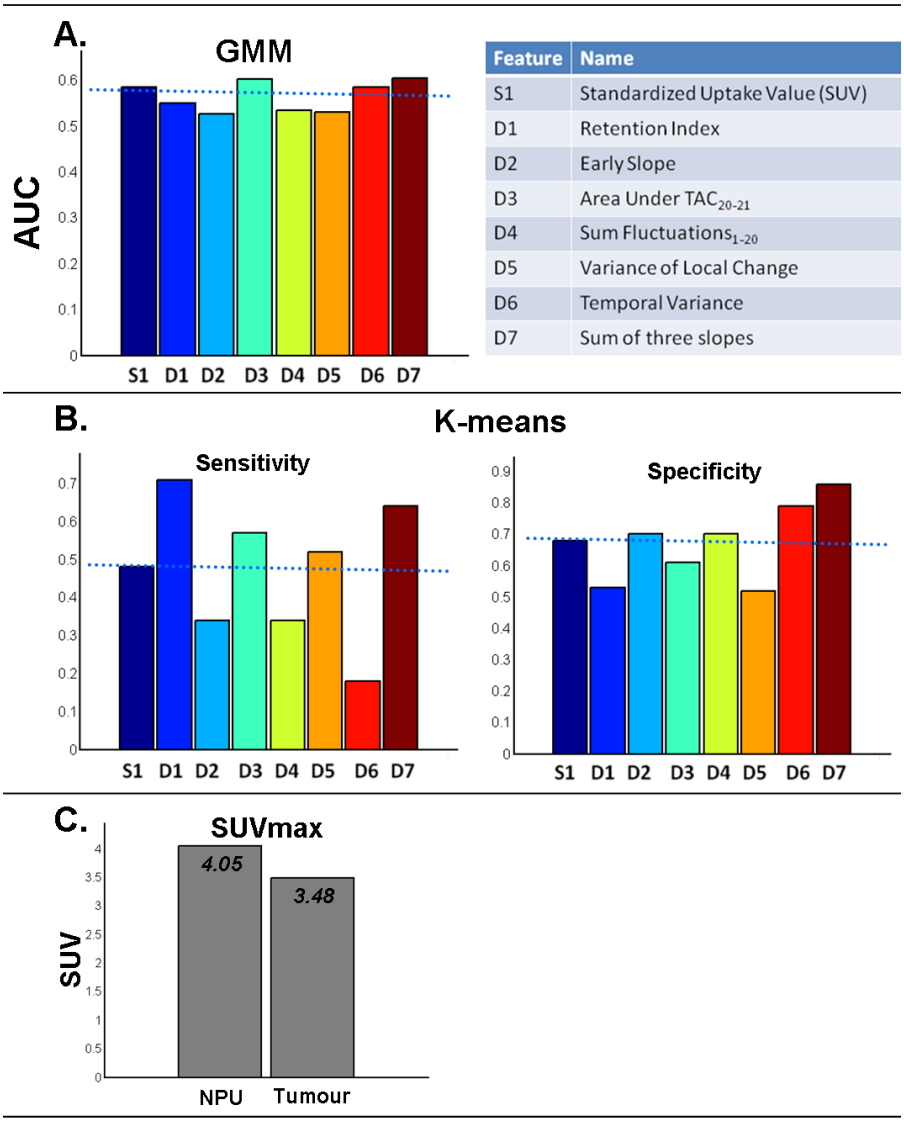


Figure 17: *Flowchart of the processes involved in the experiments of discriminating GTV from normal physiological uptake of the palatine tonsil.*



Feature	Name
S1	Standardized Uptake Value (SUV)
D1	Retention Index
D2	Early Slope
D3	Area Under TAC_{20-21}
D4	Sum $Fluctuations_{1-20}$
D5	Variance of Local Change
D6	Temporal Variance
D7	Sum of three slopes

Figure 18: Results of the experiments of discriminating GTV from Normal physiological uptake (NPU). A: Results in the form of ROC area under curve of the Gaussian Mixture Model clustering. The first navy blue coloured bar shows the ability of static parameter S1 (SUV) to successfully classify two classes while other bars show the ability of the dynamic features. B. Showing the performance of K-means classification algorithm in terms of sensitivity and specificity for its ability of the developed features to discriminate between two classes (GTV and physiological uptake in palatine tonsil). The horizontal dotted blue line in both A & B indicates level of static parameter S1 (SUV) to successfully classify two classes. C: SUV_{max} values for the normal physiological uptake (NPU) and tumour VOIs.

Results of the GTV and normal physiological uptake experiments show that both of the algorithms struggle a great deal in order to discriminate between normal physiological uptakes and FDG uptakes from the palatine tonsil carcinoma. Both of these regions show similar patterns of uptake and washout that the dynamic features vectors obtained from TACs are incapable of differentiating between these two regions. Hence, clinical findings and histopathology remains very important in these circumstances. Resultantly, based on our obtained results, from only one patient's data, we can deduce that hypothesis $\mathcal{H}3$ cannot be accepted.

5.5 Correlation Analysis

As already mentioned in the section 4 we also seek to find correlation between mean dynamic feature of the tumour VOI and complimentary data obtained independently of the dynamic PET. We seek correlation between dynamic feature and the following complimentary data,

a) T stage of the patient
b) Histopathological grade of SCC
c) HPV status of the patient

The correlations were represented by P-value calculated using Wilcoxon rank-sum test, and were also compared for dynamic features and SUVs.

5.5.1 Correlation between T-stage and Dynamic Feature (D1)

The following experiment addresses hypothesis $\mathcal{H}4$ by investigating correlation between T stage of the patients and dynamic features. We first evaluated the ability of the mean SUV of the entire tumour VOI to discriminate between the patients of different stages and later used our dynamic features. Figure 19 show the results of such correlation.

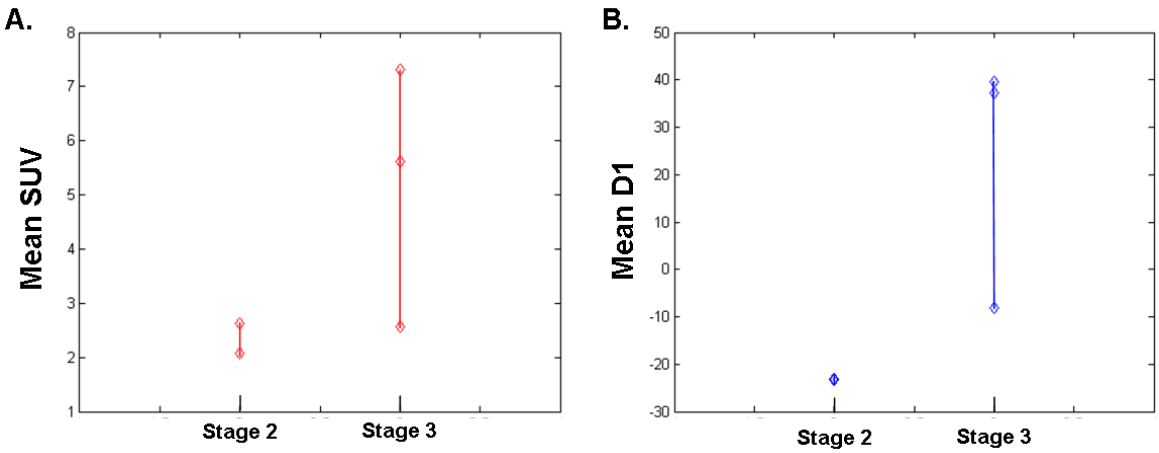


Figure 19*: Relationship between T-stage and both static as well as dynamic features; A: Mean SUV and T-stage of the patient (P=0.4 between stages 2 and stage 3 patients); B Mean Dynamic feature D1 (retention index) and T-stage of the patient (P=0.2 between stages 2 and stage 3 patients).*

Results of these correlations provide evidence in favour of the proposed hypothesis $\mathcal{H}4$ and the hypothesis can be informally accepted as the P value obtained using dynamic feature (P=0.2) is closer to our significance cut-off value of 0.05 as compared to P value obtained using SUV (P=0.4); however, more data are needed to perform a robust, conclusive hypothesis test.

5.5.2 Correlation between Histopathology SCC grade and Dynamic Feature (D5)

To address hypothesis $\mathcal{H}5$, we performed an experiment to evaluate correlation between histopathology SCC grade and dynamic features. We first evaluated the ability of the mean SUV of the entire tumour VOI to discriminate between tumours of different grades and later used our dynamic features. We found that our dynamic feature D5 provides negative correlation with tumour grade. Figure 20 show the results of such correlation.

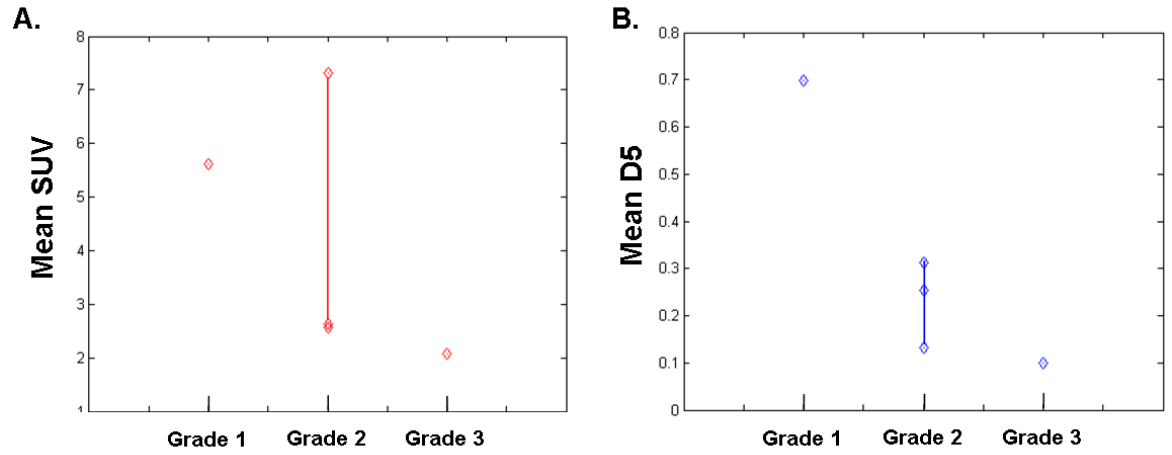


Figure 20*: Relationship between histopathology SCC grade and both static (SUV) as well as dynamic feature (D5). A: Mean SUV and histopathology SCC grade (P=1 (grade 1 & 2 or 1&3) and P=0.5 (grade 2 &3); B: Mean Dynamic feature D5 and histopathology SCC grade. (P=0.5 (grade 1 & 2 or 2&3) and P=1 (grade 2 &3)*

These results clearly show that dynamic feature (D5) can better discriminate between tumour of different grades as compared to the static parameter SUV and provide evidence in favour of hypothesis ℋ5. This differentiation maybe because of the reason that tumours of different grades possibly have different glucose metabolic rates which dynamic feature 5 (D5) was successfully able to extract from TAC and effectively exploit in this analysis.

5.5.3 Correlation between HPV status and Dynamic Feature (D5)

The final experiment looks at the correlation between HPV status and dynamic features (hypothesis ℋ6). We first evaluated the ability of the mean SUV of the entire tumour VOI to discriminate between tumours of different HPV status and later used our dynamic features. We found that our dynamic feature 5 also correlated with different HPV status. Figure 21 show the results, which suggest accepting our proposed hypothesis ℋ6.

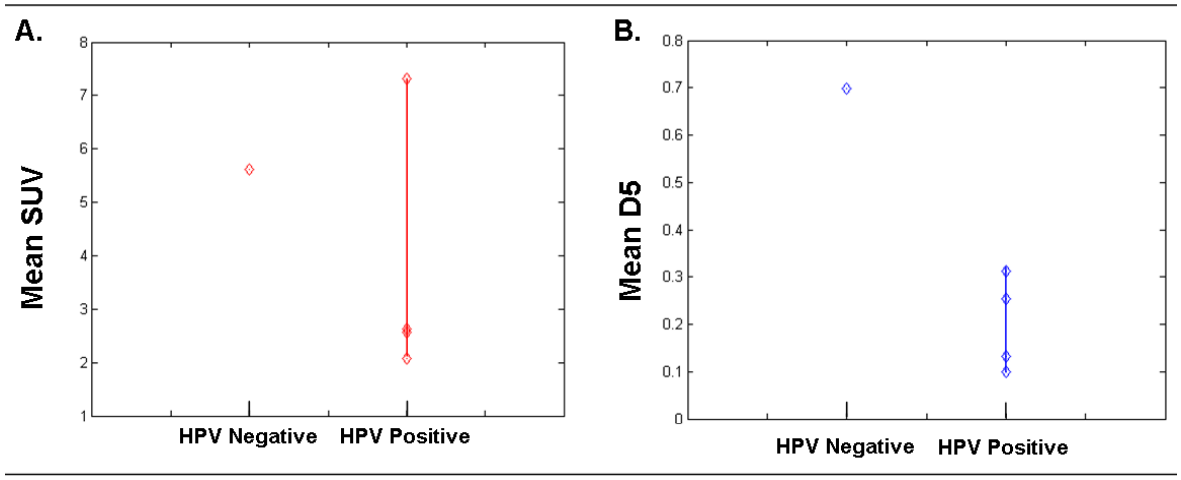


Figure 21*: Relationship between HPV status and both static (SUV) as well as dynamic feature (D5); A: Mean SUV and HPV status (P=0.8); B Mean Dynamic feature D5 and HPV status (P=0.4).*

However, it should be noticed that sample size in all of the correlation analyses was quite small as there were only five patients in our study. These results only provide an initiation point for further analysis and can prove to be useful if validated by analyzing more patients' tumour VOIs using the same analysis methodology.

6. Discussion

We have evaluated and compared 19 VOIs belonging to tumour, inflammation, healthy tissue and normal physiological uptake regions, using 2 classification algorithms and utilising 1 static (SUV) and 7 dynamic features (from TAC). In total 5774 voxels belonging to these 19 VOIs were assessed using algorithms and their intrinsic uptake-washout pattern over the 22 PET frames was exploited for differentiating metabolically distinct areas.

We present the dynamic features computed from the time activity curves of the voxels; each of these features posses different biological meanings for the reason of radiotracer uptake-washout pattern and some of the features do not depend upon the final accumulated radioactivity concentration and exploit the information of the magnitude of flux in and out of the radiotracer or take into account total retention of radiotracer over the temporal profile.

It is observed that the retention index (D1) in all of the patients' healthy tissue VOIs are below -31 and the range is in between -52 to -31 (Table 3). Retention index for the tumours' VOIs has a wide range, from -23 to 41, and it is encouraging to observe that there is no overlap between ranges of retention indices of normal tissue and tumour ones on VOI scale; therefore we can infer that the healthy and unhealthy tissue's VOI can be differentiated based on their retention indices. Voxel-wise analyses using our clustering algorithms reveal that there are many voxels within the tumour's manually delineated contours which are showing R.I similar to that of healthy tissue. It can be seen in figure 15 (C) that the R.I is showing higher sensitivity but low specificity which possibly indicates towards the presence of some voxels similar to negative class (healthy tissue) in the dataset of positive class (tumour). The reason for this is the fact that during delineation, radiation oncologist have the tendency to over segment the contours in order to not to miss any cancerous region for radiotherapy.

R.I of the palatine tonsil tumours' VOIs (range -23 to -10) are less than that of other tumours of head and neck region (range 40 to 41) which means that tumours of the tonsils have less ability of retaining FDG within themselves (Table 3). There is still a gap in published scientific knowledge of the reasons for lower FDG retention in palatine tonsils.

We have observed different shapes of the time activity curves within the same patients which represent different metabolic processes (Figure 9) and it was also interesting to see that the shape of time activity curves of tumour VOIs of the tonsil carcinoma patients were different from the non-tonsil carcinoma patients, despite the fact that they both belong to squamous cell carcinoma (SCC) of head and neck region (figure 22). It has been widely documented that tumours show steep rising curves (Sakamoto et al., 1997); however, we noticed that while this trend is not very obvious in our palatine tonsil carcinoma patients, tumour VOI of non-tonsil carcinoma (base of tongue, oropharyngeal) do possess this attribute. Inflammation or benign tissue TACs are generally considered to have flat or downward curves (Gupta et al., 1998); however, we also noticed great variability in this trend. Some of the inflammation TACs were showing steep rising curves compared to tumour ones (Figure 9) which may be because of the activity of more GLUT-3 in the inflammatory regions (section 2.3) and lesser in the cancerous region. Mellanen et al. (1994) has shown that there is generally lower expression of GLUT-3 proteins as compared to GLUT-1 proteins in squamous cell carcinoma (SCC) of head and neck region; and moreover it also has been reported that GLUT-3 has higher affinity for the transport of glucose molecules as compared to GLUT-1 (Simpson et al., 2009).

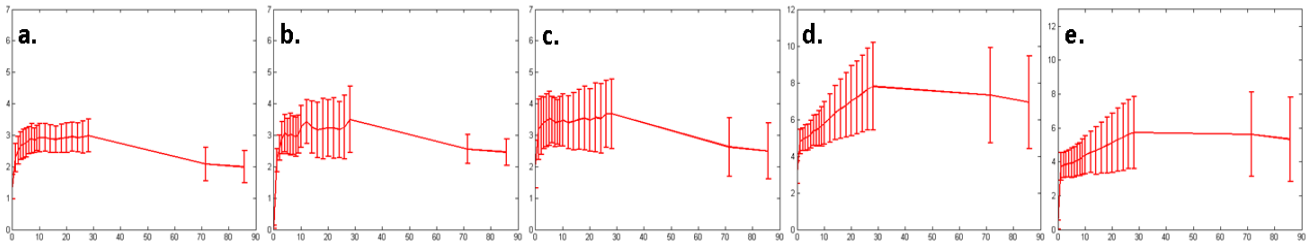


Figure 22: *Visual comparison of the shapes of the TACs of the tumour class voxels; each plot represent mean TAC with error bar indicating ± 1 standard deviation; a)*

Patient 1, palatine tonsil carcinoma; b) Patient 2, palatine tonsil carcinoma; c) Patient 3, palatine tonsil carcinoma; d) Patient 4, base of the tongue carcinoma; e) Patient 5, oropharyngeal carcinoma.

As previously mentioned, all of the inflammatory VOIs were obtained from sites of dental extraction. We also tried to find correlation between the SUVs/dynamic feature vectors and the time between dental extraction and PET scan, but we did not find any such correlation (data not shown). The type of cells present at the site of dental extraction might be playing a vital role in determining the shape of TAC (Figure 9) and the type of cells at that particular dental site also depend upon the level of difficulty of the dental procedure (e.g. healing process: platelets) and any other on-going dental ailment. Moreover, because of dental extraction, that particular site may be undergoing the various phases of soft tissue healing and all of these phases have the inclination towards certain cells types i.e. coagulation (platelets), inflammation (neutrophils, macrophages, lymphocytes), fibroplasis (fibroblasts, epithelial cells, smooth muscles, endothelial cells), remodelling (fibroblasts, macrophages) and each of the cell type have different affinity for the level and extent of expression of glucose transport proteins (GLUTs). This can possibly explain the difference in the shapes of TACs of inflammation voxels in different patients. In addition to glucose metabolism and other molecular processes, the difference between TACs within a single patient is also governed by the variability of glucose availability and blood perfusion. If comparing TACs in between different patients it might also be different because of different cardiac and respiratory cycles.

Some of the dynamic features are showing better sensitivity and specificity with K-means clustering algorithm as compared to Gaussian Mixture Model clustering algorithm; this could be explained by some computations of the dynamic features producing feature vectors which were not normally distributed and GMM algorithm sometimes struggles in classifying two groups which are not distributed normally. The Lilliefors test, which is used to check whether the data are normally distributed or not,

can be used prior to a clustering algorithm and those feature vectors which do not fulfil the criterion of being Gaussian distributed can be classified using K-means algorithm. However, in our present study we aimed to use both of the algorithms and later on assess which algorithm produce better results. The use of Lilliefors test can also be helpful if some computer based application is developed for the automated image analysis. Some of the features are also showing opposite to what stated earlier and are demonstrating better results with GMM as compared with K-means algorithm which can be because of the reason that GMM utilises EM algorithm which estimates the maximum likelihood of the data point belonging to either of the classes.

Physiological uptake at the site on palatine tonsil (patient 2) was showing the same uptake-washout pattern as compared to the tonsil carcinoma at the contra-lateral tonsil region. Dynamic features computed from TACs were unable to extract any information on the basis of which both uptake regions could be differentiated and classification algorithms also struggled to differentiate between these two classes (figure 18). Davison et al. (2010) has suggested that the SUV_{max} ratio (SUV_{max} of primary tumour/SUV_{max} of the contra-lateral tonsil) is a sensitive and specific parameter for differentiating benign metabolic activity from squamous cell carcinoma of palatine tonsil. We also utilised Davison et al. findings and calculated the SUV_{max} ratio; our region of normal physiological uptake was showing SUV_{max} of 4.08 and cancerous region SUV_{max} of 3.48 (figure 18 – C) and its ratio was 0.85 which does not correlate with Davison's findings.

Our dynamic feature D5 was able to provide some correlation between the magnitude of the feature's mean in the tumour VOI and histopathology grade of SCC. As explained in section 4.4.5, that this feature depends upon calculating the variance in the vector of local changes (in the consecutive PET frames) over the entire TAC of the particular voxel. We have seen in figure 20 that grade 1 tumours have more variance in the local changes vector, as compared to grade 2 and 3 tumours which possibly point out toward more flux in and out of the radiotracer in well differentiated squamous cell carcinoma as

compared to intermediate (grade 2) and poorly differentiated squamous cell carcinoma (SCC).

In our scanning protocol we did not measure the blood glucose level of the patients prior to scanning; routine procedure at our hospital is to measure the glucose level of only those patients who have the history of diabetes. All patients are asked to fast four hours prior to scanning and it is considered that the patient will have blood glucose level within normal limits. However, we noticed surprisingly higher SUVs in patient 4 (Table 3) which can be a result of normal blood glucose level irregularity; Busing et al. (2012) has also shown that blood glucose level (BGL) of the patients affect the bio-distribution of the FDG in muscular tissue and brain; elevated BGL is associated with reduced cerebral FDG uptake and increased musculoskeletal uptake. However, they have also reported an interesting finding that BGL does not have any significant affect on the tumour uptake of FDG. For the future research we recommend that it will be valuable to measure blood glucose level in all of the patients irrespective of their diabetic history and, using some data analysis technique, SUV images may be further normalised.

7. Conclusions

The total number of patients included in our study was relatively small and therefore the obtained results lack statistical power. Although some of the methods used in the experiments appear to be of promise in distinguishing either tumour from inflammation, or tumour from normal tissue; moreover, correlations between novel dynamic features and clinical information also demonstrate some encouraging outcomes. These results if developed further, using more patients' data and multidimensional feature space of classification algorithms, may prove to be useful in developing real world applications for diagnosis, detection and contouring. Those dynamic features which have shown promising results for differentiating tumour from inflammation may prove to be useful in reducing false positive detection rates and those features that provide positive correlations between tumours of different grades, stage and HPV can be useful in diagnosis as well as understanding the disease itself.

The main study outcomes are,

- Dynamic features (D1, D3, and D6) provide some evidence of superior ability of differentiating tumour voxels from inflammatory voxels as compared to SUV which may assist physicians during PET image interpretations.
- NPU and tumour of the palatine tonsil, within the same patient, may show similar radioactivity uptake washout patterns evident in TAC; therefore histopathological analyses remain important in these circumstances.
- Dynamic feature D1 and D5 correlate with clinical parameters obtained independently of PET imaging (stage and grade - HPV status of the tumour, respectively).

In conclusion, the present study has definitely opened some new doors for future research and highlighted the areas which need to be understood for further scientific development within this domain. Moreover, after further investigations, these methods can be applied in clinical setting and building up some computer based application for diagnosis, detection and contouring.

Acknowledgments

I would like to acknowledgment my parents, ***Talib Mahmood*** *&* ***Malahat Sheikh****, for their tremendous support and love, without their active motivation I would never have been able to come to Finland for studies.*

My brilliant supervisor, ***Tony Shepherd****, who kept on guiding & teaching new methods and always remained there whenever I needed his support. Despite his other work related commitments he always took time out for me and kept my morale high even in my darkest times. He invested a lot of his time in me and never objected to my stupid ideas and silly questions.*

Prof. Heikki Minn *and* ***Antti Silvoniemi*** *at Turku University Hospital (TYKS) who always provided their expert opinions and encouraged throughout the whole process of master's thesis work.*

Coordinators of Master's degree programme in Biomedical Imaging ***Maritta Löytömäki*** *and* ***Eeva Rainio*** *who were all the time very active in resolving problems.*

Prof. John Eriksson *and* ***Prof. Pekka Hänninen*** *for their active support, motivation and guidance.*

My old colleagues, supervisors and teachers, particularly ***Najam Uddin****,* ***Zia Faruqui****,* ***Ahmed Murtaza****,* ***Memoona Mian****,* ***Muhammad Abdul Raffaye****,* ***Muhammad Ali Khan*** *and* ***Shama Murad*** *at* ***Shaukat Khanum Memorial Cancer Hospital and Research Centre, Lahore, Pakistan*** *for developing scientific insight in me during my years spent there.*

And finally, my siblings, ***Arooj****,* ***Adiya*** *&* ***Hamza,*** *and friends,* ***Shariq,*** *&* ***Fahad*** *who regularly remained in touch with me and also made me happy by sharing their gossips.*

References

Adams, S., R.P. Baum, T. Stuckensen. 1998. Prospective comparison of FDG PET with conventional imaging modalities (CT, MRI, US) in lymph node staging of head and neck cancer. *Eur J Nucl Med*. 25:1255– 1260.

Altman, D.G., and J.M. Bland. 1994. Diagnostic tests 1: sensitivity and specificity. Department of Public Health Sciences, St George's Hospital Medical School, London.

Bailey, B.J., J.T. Johnson, S.D. Newlands. 2006. Head & Neck Surgery: Otolaryngology. Lippincott Williams & Wilkins, New York, 199 pp.

Bakheet, S.M., and J. Pow. 1998. Benign causes of ^{18}-FDG uptake on whole body imaging. *Semin Nucl Med*. 28:352-358

Basu, S., J. Kung, M. Houseni, H. Zhuang, G. Tidmarsh, and A. Alvi. 2009. Temporal profile of fluorodeoxyglucose uptake in malignant lesions and normal organs over extended time periods in patients with lungs carcinoma: implications for its utilization in assessing malignant lesions. *Q J Nucl Med Mol Im*. 53(1):9-19.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer: Verlag

Boellaard, R. 2009. Standards for PET image acquisition and quantitative data analysis. *J. Nucl. Med*. 50:11-20.

Boellaard, R., M.J. O'Doherty, W.A. Weber, F.M. Mottaghy, M.N. Lonsdale. 2010. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging*. 37:181–200.

Büsing, K.A., S.O. Schönberg, J. Brade, and K. Wasser. 2012. Impact of blood glucose, diabetes, insulin, and obesity on standardized uptake values in tumors and healthy organs on F-FDG PET/CT. *Nucl Med Biol*. 40(2):206-213.

Calvo, M. B., A. Figueroa, E.G. Pulido, R.G. Campelo, and L.A. Aparicio. 2010. Potential role of sugar transporters in cancer and their relationship with anticancer therapy. *Int J Endocrinol*. 2010:1-14.

Carvalho, A., I. Nishimoto, J. Califano, and L. Kowalski, 2005. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. *Int J Cancer*. 114: 806-816.

Cherry, S.R. 2006. PET: Physics, Instrumentation, and Scanners. Springer, Los Angeles.

Chiu, C.F., Y.Y. Lin, W.H. Hsu, C.Y. Chen, J.J. Yeh, and C.H. Kao. 2012. Shorter-time dual-phase FDG PET/CT in characterizing solid or ground-glass nodules based on surgical results. *Clin Imag*. 1-6.

Davison, J.M., A. Ozonoff, H. M. Imsande, G.A. Grillone, and R.M. Subramaniam. 2010. Squamous cell carcinoma of the palatine tonsils: FDG standardized uptake value ratio as a biomarker to differentiate tonsillar carcinoma from physiologic uptake. *Radiology*. 255(2):578-585

Delbeke, D., W.H. Martin, J.A. Patton, and M.P. Sandler. 2002. Practical FDG Imaging: A Teaching File. Springer, Nashville.

Dempster, A. P., N.M. Laird, and D.B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc*. 39(1):1–38.

Epstein, J.B., and P.S. Moore. 2001. Periodontal disease and periodontal management in patients with cancer. *Oral Oncol*. 37:613– 619.

Frisch, M., H. Hjalgrim., A.B. Jaeger. 2000. Changing patterns of tonsillar squamous cell carcinoma in the United States. *Cancer Causes Control*. 11:489–95.

GE Healthcare. 2011. *Clinical Implementation of VUE Point FX™ white paper*. Available online: http://www3.gehealthcare.co.uk/~/media/Downloads/uk/Education/PET%20white%20papers/MI_EMEA_VUE_Point_FX_White_Paper_PDF_092009_DOC0659348.pdf?Parent={B66C9E27-1C45-4F6B-BE27-D2351D449B19} (Accessed: April 26, 2013)

Gillies, R.J., I. Robey, and R.A. Gatenby. 2008. Causes and consequences of increased glucose metabolism of cancers. *J Nucl Med*. 49(6):24–42.

Gordin, A., A. Golz, G. Keidar, M. Daitzchman, B.R. Shalom, and O. Israel. 2007. The role of FDG-PET/CT imaging in head and neck malignant conditions: Impact on diagnostic accuracy and patient care. *Otolaryng Head Neck*. 137:130-137

Górriz, J.M., A. Lassl, J. Ramírez, D.S. Gonzalez, C.G. Puntonet , and E.W. Lang. 2009. Automatic selection of ROIs in functional imaging using Gaussian mixture models. *Neurosci Lett* 460:108–111.

Guo, Q., J. Shao, V.F. Ruiz. 2009. Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms. *Int J Comput Assist Radiol Surg*. 4(1):11-25

Gupta, N., H. Gill, G. Graeber, H. Bishop, J. Hurst, and T. Stephens, 1998. Dynamic positron emission tomography with F-18 fluorodeoxyglucose imaging in differentiation of benign from malignant lung/mediastinal lesions. *Chest*. 114:1105-1111

Haerle, S. K., D. Schmid, N. Ahmad, T. Hany, and S. Stoeckli. 2011. The value of $^{(18)}$ F-FDG PET/CT for the detection of distant metastases in high-risk patients with head and neck squamous cell carcinoma. *Oral Oncol*. 47(7):653–659.

Hanley, A, J., and J.B. McNeil. 1982. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *J Radiol*. 143(1):29-36.

Hall, E.L., R.P. Kruger, S.J. Dwyer, D.L. Hall, R.W. Mclaren, G.S. Lodwick. 1971. A survey of preprocessing and feature extraction techniques for radiographic images**.** *IEEE Trans Comput*. 20:1032–44

Hemminki, K., C. Dong, and M. Frish. 2000. Tonsillar and other upper aero-digestive tract cancers among cervical cancer patients and their husbands. *Eur J Cancer Prev*. 9:433–437.

Higashi, T., T. Saga, Y. Nakamoto, T. Ishimori, M.H. Mamede, M. Wada, R. Doi, R. Hosotani, M. Imamura, and J. Konishi. 2002. Relationship between retention index in dual-phase ^{18}F-FDG PET, and hexokinase-II and glucose transporter-1 expression in pancreatic cancer *J Nucl Med* 43:173–180.

Hubner, K.F., J.A. Thie, G.T. Smith, A.C. Chan, P.S. Fernandez, and J.M. McCoy. 2000. Clinical utility of FDG-PET in detecting head and neck tumors: a comparison of diagnostic methods and modalities. *Clin Positron Imaging*. 3:7-16.

Hustinx, R., R.J. Smith, F. Benard, D.I. Rosenthal, M. Machtay, L.A. Farber, et al. 1999. Dual time point fluorine-18 fluorodeoxyglucose positron emission tomography: a potential method to differentiate malignancy from inflammation and normal tissue in the head and neck. *Eur J Nucl Med*. 10:1345-1348.

Ion, A.L. 2011. Rough sets and Gaussian mixture model in medical image diagnosis. *Annals of the University of Craiova, mathematics and computer science series*. 38(4):50-62.

Jadvar, H., A. Alavi, and S.S. Gambhir. 2009. ^{18}F-FDG uptake in lung, breast, and colon cancers: Molecular biology correlates and disease characterization. *J Nucl Med*. 50:1820-1827.

Janssen, M.H.M., H.J. Aerts, L. Michel, G. Bosmans, J.A. Lee, J. Buijsen, D.D Ruysscher, P. Lambin, G. Lammering, and A.L. Dekker. 2009. Tumor delineation based on time–activity curve differences assessed with dynamic fluorodeoxyglucose positron emission tomography–computed tomography in rectal cancer patients. *Int. J. Radiation Oncology Biol Phys*. 73(2): 456–465.

Kresnik, K., P. Mikosch, H.J. Gallowitsch. 2001. Evaluation of head and neck cancer with ^{18}F-FDG PET: a comparison with conventional methods. *Eur J Nucl Med*. 28:816 –821.

Laffon, E., H.D. Clermont, and R. Marthan. 2011. A method of adjusting SUV for injection-acquisition time differences in F-FDG PET Imaging. *Eur Radiol*. 21(11):2417-2424.

LeVine, H. 2010. Medical Imaging. Connecticut: Greenwood Publishing Group.

Lewin, F., S.E. Norell, H. Johansson. 1998. Smoking tobacco, oral snuff, and alcohol in the etiology of squamous cell carcinoma of the head and neck. A population-based case-referent study in Sweden. *Cancer*. 82:1367–75.

Lin, E.C., and A. Alavi. 2009. PET and PET/CT: A Clinical Guide. New York: Thieme Medical Publisher.

Lodge, M.A., J.D. Lucas, P.K. Marsden, B.F. Cronin, M.J. O'Doherty, and M.A. Smith. 1999. A PET study of ^{18}FDG uptake in soft tissue masses. *Eur J Nucl Med*. 26:22–30.

Macheda, M.L., S. Rogers., J.D. Bets. 2005. Molecular and cellular regulation of glucose transport (GLUT) proteins in cancer. *J Cell Physiol*. 202:654–662.

Mason, S.J., and N.E. Graham. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation *Q.J.R. Meteorol. Soc*. 128: 2145–2166.

Matlab Central – File Exchange Server. 2013. *The MathWorks, Inc*. Massachusetts, United States Available online http://www.mathworks.se/matlabcentral/fileexchange/ (Accessed: August, 2012)

Mellanen, P., H. Minn, R. Grenmanan, and P. Harkonen. 1994. Expression of glucose transporters in head-and-neck tumors. *Int. J. Cancer*. 56:622-629.

Metz, C.E. 2006. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J Am Coll Radiol*. 3:413-422.

Miles, K.A., and R.E. Williams. 2008. Warburg revisited: imaging tumour blood flow and metabolism. *Cancer Imaging*. 8:81-86.

Minn H., S. Suilamo, and J. Seppala. 2010. Impact of PET/CT on planning radiotherapy in head and neck cancer. *Q J Nucl Med Mol Imaging*. 54:521-532.

Morris, E.D., C.J. Endres, K.C. Schmidt, B. T. Christian, RF. Muzic, R.E. Fisher. 2004. Kinetic Modeling in Positron Emission Tomography. *Emission Tomography*. 499-540

Murphy, K. P. 2007. *Performance evaluation of binary classifiers*. Availiable online: http://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall07/reading/rocHandout.pdf (Accessed: February 13, 2013).

Muzi. M., O' Sullivan, D.A. Mankoff, R.K. Doot, L.A. Pierce, B.F. Kurland, H.M. Linden, and P.E. Kinahan. 2012. Quantitative assessment of dynamic PET imaging data in cancer imaging. *Magn Reson Imaging*. 30(9):1203-1215.

Nahmias, C. and L.M. Wahl. 2008. Reproducibility of standardized uptake value measurements determined by ^{18}F-FDG PET in malignant tumors. *J Nucl Med*. 49(11):1804-1808.

Ng, H.P., S.H. Ong, K.W.C Foong, P.S. Goh, and W.L. Nowinski. 2006. Medical image segmentation using K-means clustering and improved watershed algorithm. *Image Analysis and Interpretation, IEEE Southwest Symposium*. 1:61-65.

Oyelade, O. J., O.O. Oladipupo, and I.C. Obagbuwa. 2010. Application of K-means Clustering algorithm for prediction of Students' Academic Performance. *Int J Comput Inf Sc*. 7(1):292-295

Park, S., J. Goo, and C. Jo. 2004. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol*. 5(1):11-8.

Pham, D.L., C. Xu, and J.L. Prince. 2000. A survey of current methods in medical image segmentation. *Annu Rev Biomed Eng*. 2:315–337.

Pham, D.L. and P.L. Bazin. 2008. Unsupervised tissue classification. In: *Handbook of Medical Image Processing and Analysis* (edited by Isaac Bankman), pp. 209-222. Academic Press, Massachusetts.

Plathow, C., and W.A. Weber. 2008. Tumor cell metabolism imaging. *J Nucl Med*. 49 (6):43–63.

Reynolds, D. (n.d). *Gaussian Mixture Models*. MIT Lincoln Laboratory, Lexington. Availiable online: http://www.ll.mit.edu/mission/communications/ist/publications/0802_Reynolds_Biometrics-GMM.pdf (Accessed: December 11, 2012).

Saha, G.B. 2010. *Basics of PET Imaging: Physics, Chemistry, and Regulations*. Springer, NewYork.

Sakamoto, H., Y. Nakai, and Y. Ohashi. 1997. Positron emission tomographic imaging of head and neck lesions. *Eur Arch Otorhinolaryngol*. 254:123-126.

Shepherd, T., and R. Owenius. 2012a. Gaussian process models of dynamic pet for functional volume definition in radiation oncology. *IEEE Trans Med Imaging*. 31(8):1542-56.

Shepherd, T., M. Teräs, R.R. Beichel, R. Boellaard, M. Bruynooghe, V. Dicken, M.J. Gooding, P. J. Julyan, J. A. Lee, S. Lefèvre, et al. 2012b. Comparative Study with New Accuracy Metrics for Target Volume Contouring in PET Image Guided Radiation Therapy. *IEEE Trans Med Imaging*. 31(11):2006:2024

Shreve, P.D., Y. Anzai, and R.L. Wahl. 1999. Pitfalls in oncologic diagnosis with FDG PET imaging: phsysiologic and benign variants. *Radiographics*. 19(1):61-77.

Simpson, I.A., D. Dwyer, D. Malide, K.H. Moley, A. Travis, S.J. Vannucci. 2008. The facilitative glucose transporter GLUT3: 20 years of distinction. *Am J Physiol Endocrinol Metab*. 295(2):242–253.

Sobin, L. H. 2003. TNM: Evolution and relation to other prognostic factors. *Semin. Surg. Oncol*. 21:3–7.

Su, M.C., and C.H. Chou. 2001. A modified version of the k-means algorithm with a distance based on cluster symmetry. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(6):674-680

Syrjänen, K., S. Syrjänen. 2000. Papillomavirus infections in human pathology. Wiley, London.

Syrjanen, S. 2004. HPV infections and tonsillar carcinoma. *J Clin Pathol*. 57(5):449–455.

Szablewski, L. 2013. Expression of glucose transporters in cancers. *Biochimica et Biophysica Acta*. 1835(2):164-9.

Thie, J. A. 2004. Understanding the standardized uptake value, its methods, and implications for usage. *J Nucl Med*. 45:1431-1434.

Thorwarth, D., S.M. Eschmann, J. Scheiderbauer. 2005. Kinetic analysis of dynamic 18Ffluoromisonidazole PET correlates with radiation treatment outcome in head-and-neck cancer. *BMC Cancer*. 5:152.

Wenig. B.E. and J.M. Cohen. 2009. General Principles of Head and Neck Pathology. In: *Head and Neck Cancer: A Multidisciplinary Approach* (edited by Louis B., B. Harrison, R.B. Sessions, and W.K. Hong), pp.3-17. Lippincott Williams & Wilkins: New York.

Yao, M., and P.F. Faulhaber, 2012. PET Imaging of the Head and Neck, An Issue of PET Clinics. *PET Clinics*. 7(4):345-466.

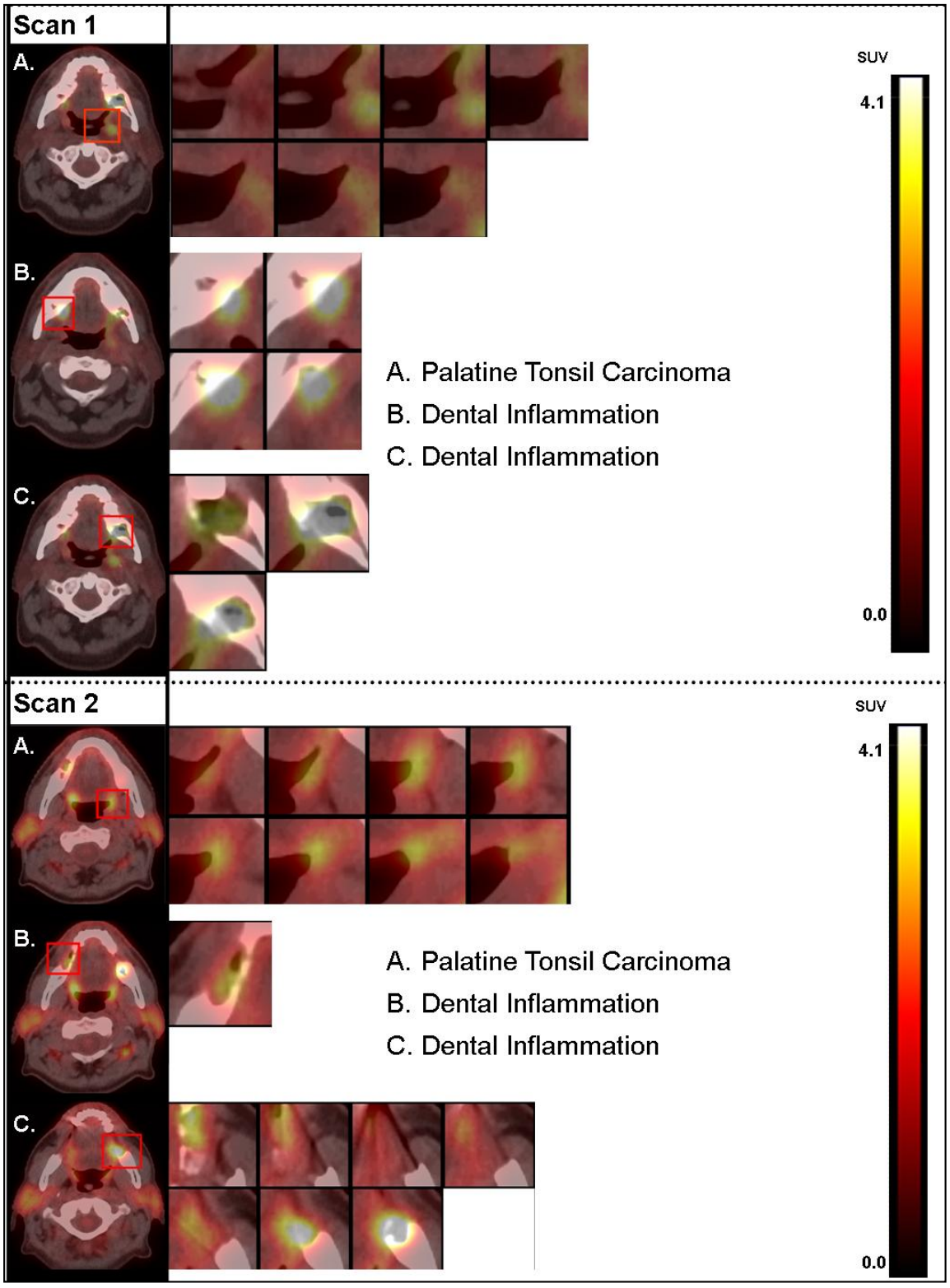
Younes, M., L.V. Lechago, J.R. Somoano. 1997. Imunohistochemical detection of glut3 in human tumors and normal tissues. *Anticancer Res*. 17(4):2747–2750.

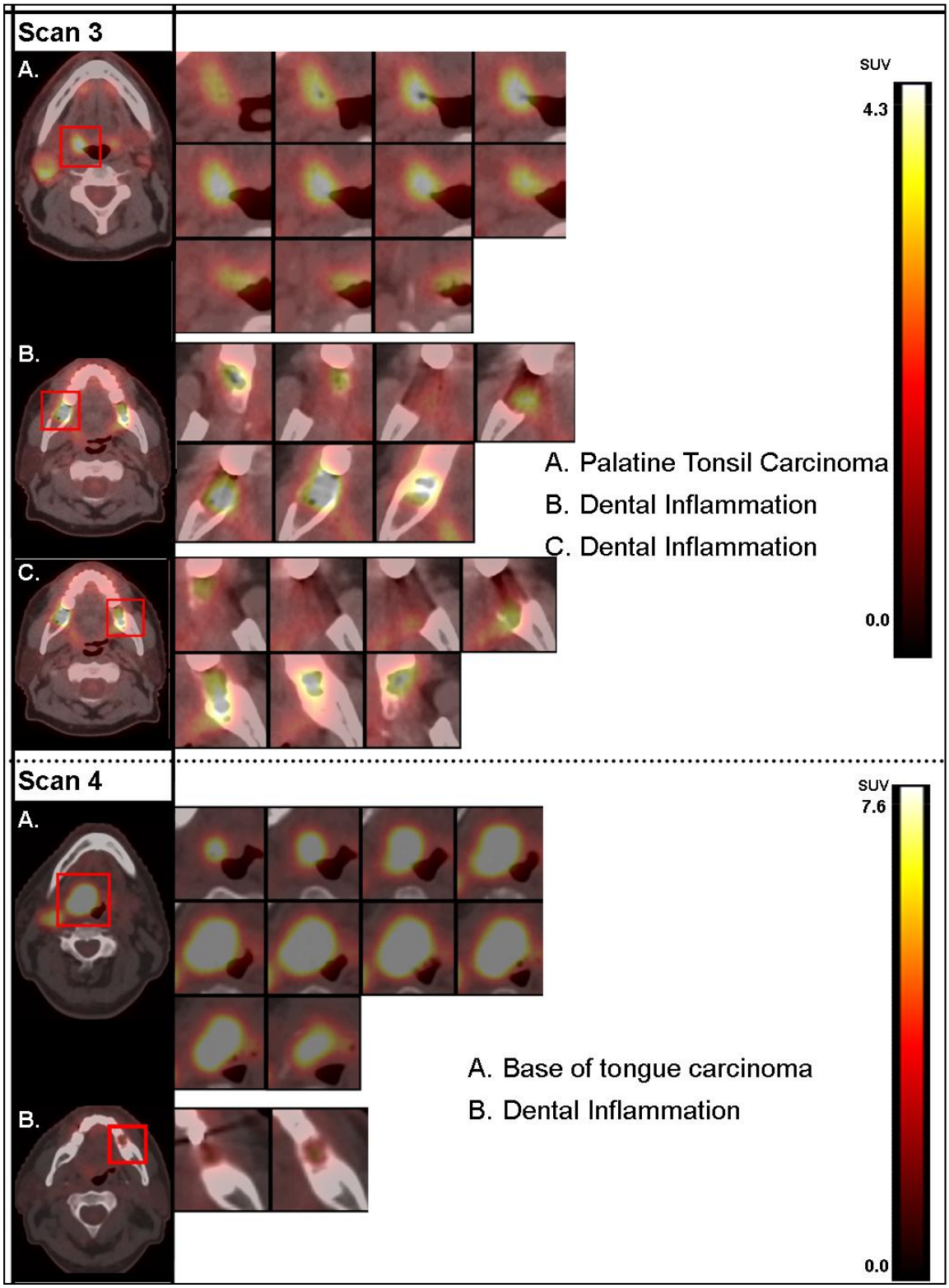
Zaidi, H., H. Vees, and M. Wissmeyer. 2009. Molecular PET/CT Imaging-Guided Radiation Therapy Treatment Planning. *Acad Radiol*. 16(9):1108-1133

Zhou, S., S. Wang, Q. Wu, J. Fan, and Q. Wang. 2008. Expression of glucose transporter-1 and -3 in the head and neck carcinoma—the correlation of the expression with the biological behaviours. *ORL J Otorhinolaryngol Relat Spec*. 70(3):189–194.

Zhuang, H., M. Pourdehnad, E.S. Lambright, A.J. Yamamoto, M. Lanuti, P. Li, P.J. Mozley, M.D. Rossman, S.M. Albelda, and A. Alavi. 2001. Dual time point ^{18}F-FDG PET imaging for differentiating malignant from inflammatory processes. *J Nucl Med*. 42:1412–1417.

Appendix A


Scan 1
A.
B.
C.
SUV
4.1
0.0
A. Palatine Tonsil Carcinoma
B. Dental Inflammation
C. Dental Inflammation
Scan 2
A.
B.
C.
SUV
4.1
0.0
A. Palatine Tonsil Carcinoma
B. Dental Inflammation
C. Dental Inflammation


Scan 3
A.
B.
C.
SUV
4.3
0.0
A. Palatine Tonsil Carcinoma
B. Dental Inflammation
C. Dental Inflammation
Scan 4
A.
B.
SUV
7.6
0.0
A. Base of tongue carcinoma
B. Dental Inflammation

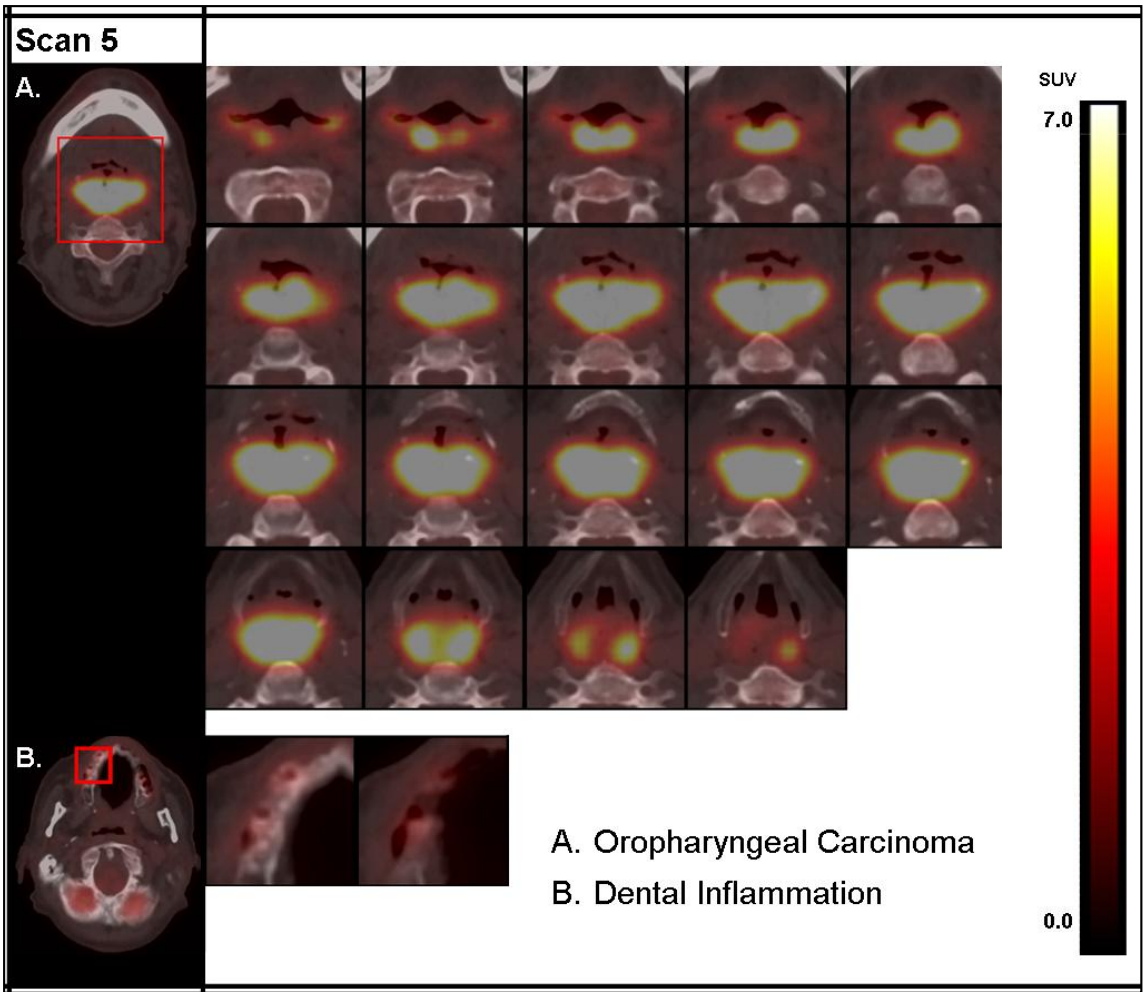


Figure Appendix A: *Showing all the sections, along the craniocaudal axis, of all those VOIs which have considerable ^{18}F-FDG uptake and were included in the study.*