survival (OS) in months was used as model output in this study. For external validation, a real-life patient cohort fulfilling the same inclusion criteria as the training set was retrospectively collected. A 3-fold cross-validation was used on the training set for feature and model selection and hyperparameter optimization. To avoid overfitting, recursive feature elimination was used to select the most performant features. Model performance was evaluated by calculating the median absolute error between the ground-truth OS and predicted OS. ***Results:*** Data from 252 and 74 patients were available for training and external validation, respectively. Mean OS [min-max] equalled 8.8 [0.1-55.6] months in the training cohort and 5.6 [0.4-22.9] months in the validation cohort. The features retained were WB-MATV, SDmax, visceral fat density, number of metastatic sites, ECOG-PS, days since diagnosis and body mass index. A support vector regressor with radial basis function kernel was selected as the final model. The median absolute error [75%, 95% percentile] of the trained model averaged over the 3 training folds was 1.1 [2.6, 6.4] months, and 1.4 [2.5-7.8] months for the external validation cohort. ***Conclusion:*** This study confirms the strong prognostic power of ^{18}F-FDG PET/CT-based biomarkers in advanced chemorefractory mCRC. By combining these biomarkers in a prognostic machine learning model, patient outcome could be predicted with good accuracy. This tool could provide objective data supporting clinical decision-making to identify the likelihood of treatment futility and prioritize the patient's quality of life. ***References:*** [1] Hendlisz A, et al. PLoS One (2015);10:1-14.[2] Hendlisz A, et al. BMJ Open (2015);5:e007189.[3] Camera S., et al. Cancers (2020);12(10):2752.

OP-746

Machine learning to diagnose obstructive coronary artery disease using calcium scoring, PET imaging and clinical data

S. Koenders[1,2], ***J. A. van Dalen***[3], *R. J. Metselaar[1], B. Vendel[1], D. J. Slotman[4], M. Mouden[5], C. H. Slump[2], J. D. van Dijk[1];*
[1]Nuclear Medicine, Isala Hospital, Zwolle, NETHERLANDS, [2]Technical Medical Centre, University of Twente, Enschede, NETHERLANDS, [3]Medical Physics, Isala Hospital, Zwolle, NETHERLANDS, [4]Radiology, Isala Hospital, Zwolle, NETHERLANDS, [5]Cardiology, Isala Hospital, Zwolle, NETHERLANDS.

Aim/Introduction: Accurate risk stratification using non-invasive diagnostic imaging in patients with suspected stable coronary artery disease (CAD) is essential for choosing an appropriate treatment strategy. Yet the human ability to interpret and integrate all available data into one post-test likelihood is not straightforward. Artificial intelligence can help to analyse all available data. Our aim was to develop and validate a machine learning (ML) model to diagnose obstructive CAD in patients, based on the available clinical data, medication and CT-based coronary artery calcium scoring (CACS) and Rubidium-82 (Rb-82) PET data. In addition, we compared the performance of this ML model to that of expert physicians. ***Materials and Methods:*** We retrospectively included 1007 patients who underwent CACS and a Rb-82 PET scan without a prior history of CAD. The entire dataset was split 4:1 into a training and test dataset. An XGBoost model was developed and optimized with the training dataset, using hyperparameter optimization via grid search in combination with 5-fold stratified cross-validation. The area under the receiver operating curve (AUC) was computed for the XGBoost model for both datasets. Next, the test dataset was used to compare the accuracy, sensitivity and specificity of expert readers to the XGBoost model. The primary endpoint was obstructive CAD on invasive coronary angiography (ICA). ***Results:*** Of the included 1007 patients, 111 (11%) patients were classified as having obstructive CAD during follow-up. The median follow-up time was 1.8 years. The minimum follow-up time was 1 year and maximum follow-up time was 2.7 years. The most important predictors for the ML model consisted of CACS and PET-derived features. In addition, medications were found to hold little to non-predictive value (F-score ≤ 1). Receiver operating curve (ROC) analysis showed an AUC of 0.92 and 0.89 for the training and test dataset, respectively. The expert readers achieved an accuracy of 88%, a sensitivity of 69% and a specificity of 90%. The performance of the ML model did not differ to expert readers in accuracy (89%), sensitivity (68%) and specificity (92%, p≥0.03). ***Conclusion:*** We have developed and validated a ML model to diagnose obstructive CAD using CACS, Rb-82 PET imaging and clinical data. The ML model resulted in a similar diagnostic performance as compared to expert readers. This tool may be deployed as risk stratification tool and might help physicians in training. This study showed that utilization of ML is promising in the diagnosis of obstructive CAD.

OP-747

Assessment of the potential of convolutional neuronal networks in the differential diagnosis of Parkinson's disease based on brain imaging [^{123}I]FP-CIT SPECT

N. Valador[1,2], *F. P. M. Oliveira[1], P. M. S. Ferreira[1], L. Vieira[3,4], D. C. Costa[1];*
[1]Champalimaud Foundation, Lisbon, PORTUGAL, [2]Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Lisbon, PORTUGAL, [3]Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Lisbon, PORTUGAL, [4]Health and Technology Research Center, Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Lisbon, PORTUGAL.

Aim/Introduction: To evaluate the potential of convolutional neural networks (CNN) in the differential diagnosis of Parkinson's disease (PD) based on [^{123}I]FP-CIT single-photon emission computed tomography (SPECT) images, compared to other machine learning-based classifiers. ***Materials and Methods:*** This work included 806 [^{123}I]FP-CIT SPECT brain images (208 health controls and 598 with PD). Data were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For each subject, only the first scan was considered (baseline or screening). The protocol of image acquisition and pre-processing is available at http://www.ppmi-info.org/study-design/research-documents-and-sops/. CNN was compared against k-nearest neighbour (kNN), logistic regression (LG), decision trees (DT), support vector machines (SVM) and artificial neural networks (ANN) classifiers. The CNN classifier was trained with 2-dimensional image patches (dimensions: 88 mm x 82 mm) containing the striatal region, extracted from the head superior-inferior maximum intensity projection. The remaining classifiers were trained with five features extracted from 3-dimensional striatal region: caudate binding potential, putamen binding potential, putamen to caudate ratio, volume of the striatal region with "normal uptake", and major axis of that region. The minimum values extracted from each cerebral hemisphere were used. The split ratio of the dataset was 75:25 (75% for training and 25% for testing). Each of the five features was also considered individually to assess its potential for classification in terms of performance (accuracy, sensitivity, and specificity). ***Results:*** In the test dataset, accuracy, sensitivity, and specificity of the CNN were 96%, 98%, and 91%, respectively. This finding was very similar to what we obtained with the other classifiers (kNN: 95%, 99%, 85%; LG: 94%, 97%, 86%, DT: 94%, 97%, 84%, SVM: 94%, 98%, 88% and ANN: 94%, 97%, 86%). The accuracy differences are not statistically significant (Cochran Q test, p = 0.592). Individually, the feature that best differentiate PD from normal scans was the putamen binding potential with 93% accuracy, 93% sensitivity and 94% specificity in the test dataset, based on the optimal cutoff (1.716) that maximizes Younden's coefficient in the training dataset. ***Conclusion:*** CNN classifier proved to be as robust and accurate as the other classifiers frequently used in the type of problems as in this work, with the great advantage of using images as direct input. All machine learning-based classifiers tested are robust and very accurate in the classification of brain [^{123}I]FP-CIT SPECT scans. Standard visual clinical evaluation should be complemented with quantification classification used also as a training tool.

OP-748

Machine Learning-based classification using ^{18}F-FDG PET-derived quantitative parameters in predicting Endometrial Cancer aggressiveness

C. Bezzi[1,2], *P. Mapelli*[1,2], *G. Mathoux*[3], *L. Monaco*[3], *S. Ghezzo*[1,2], *A. Bergamini*[1,4], *F. Fallanca*[2], *A. M. Samanes Gajate*[2], *F. Vasta*[1,4], *G. Candotti*[1,4], *G. L. Taccagni*[5], *G. Mangili*[4], *L. Gianolli*[2], *M. Picchio*[1,2]; [1]*Vita-Salute San Raffaele University, Milan, ITALY,* [2]*Nuclear Medicine Department, IRCCS San Raffaele Scientific Institute, Milan, ITALY,* [3]*University of Milano-Bicocca, Milan, ITALY,* [4]*Unit of Obstetrics and Gynaecology, IRCCS San Raffaele Scientific Institute, Milan, ITALY,* [5]*Pathology Unit, IRCCS San Raffaele Scientific Institute, Milan, ITALY.*

Aim/Introduction: Despite providing valuable information for optimal treatment planning, features of endometrial cancer (EC) aggressiveness are currently assessable only after surgery. This study aims at investigating the role of ^{18}F-FDG PET in predicting EC pathological prognostic factors in patients scanned for preoperative staging, also evaluating the potential additional value of machine learning (ML) algorithms alongside conventional statistics. ***Materials and Methods:*** Retrospective monocentric study including 123 patients with histologically proven EC who underwent ^{18}F-FDG PET (2009-2021) for preoperative staging. Maximum standardized uptake value (SUVmax), SUVmean, metabolic tumour volume (MTV) and total lesion glycolysis (TLG) PET parameters were derived on the primary lesion. Patient's age and BMI were collected. Features of EC aggressiveness, including histotype, deep myometrial invasion (MI), risk group, lymph-nodal involvement (LN), and p53 expression were retrieved from histological samples and, based on data availability, different cohorts were generated. Age, BMI and PET parameters were investigated through Mann-Whitney U test. Each cohort was split into a train (70%) and a validation (30%) set. The training set was analysed through receiver operating characteristic (ROC) curves. For each parameter demonstrating a significant, predictive AUC, optimal cut-off was computed (Youden Index method), and tested to classify patients of the validation set. Finally, parameters identified as predictive were combined as inputs into ML models (Random Forest Classifiers-RFCs), that were trained and validated on the same train and validation sets, respectively. Accuracy (AC), sensitivity (SN), specificity (SP), and positive and negative predictive values (PPV, NPV) were computed from both ROC curve-derived cut-offs and ML models. ***Results:*** Of the 123 EC patients overall included in the study, 85/123 presented an endometrioid histotype (vs. non-endometrioid), 53/115 presented MI >50%, 76/119 were grouped as high-intermediate/high risk (vs. low/intermediate), 14/90 presented LN, and 37/51 had p53 overexpression. ^{18}F-FDG PET detected the primary tumour in 96,7% patients. Prediction metrics obtained on the validation set from both single parameters' cut-offs and ML models are reported in Tables 1,2,3,4. Considering EC histotype, none of the investigated parameters showed significant results. ***Conclusion:*** This study represents one of the first investigations on the role of ^{18}F-FDG PET parameters in assessing EC aggressiveness. Overall, compared to the potential of single PET parameters, the employment of ML models revealed to be a valuable approach to increase performances in predicting EC pathological prognostic factors. Moreover, compared to known clinical factors (age, BMI), PET parameters demonstrated higher predictive abilities.

OP-749

Kinetic filtering and deep learning for the automatic detection and quantification of primary brain tumors using dynamic ^{18}F-FET PET imaging

M. Rahimpour[1], *R. Boellaard*[2], *W. Deckers*[3], *K. Goffin*[3], *M. Koole*[1]; [1]*KU Leuven, Leuven, BELGIUM,* [2]*Amsterdam UMC, Amsterdam, NETHERLANDS,* [3]*UZ Leuven, Leuven, BELGIUM.*

Aim/Introduction: Dynamic ^{18}F-fluoroethyl-L-tyrosine (^{18}F-FET) PET imaging has a proven added value for the clinical management of gliomas. An important quantitative parameter for lesion characterization is the metabolic index (MI) defined as the ratio of the maximal lesion over average background uptake. Generally, manual delineations of the lesion and a crescent-shaped background region in the non-affected hemisphere are performed to estimate MI. The aim was to develop an automatic approach to detect primary brain lesions and estimate the corresponding MI from dynamic ^{18}F-FET PET. For this purpose, kinetic filtering (KF) and convolutional neural network (CNN) approach were considered and evaluated. ***Materials and Methods:*** 84 patients who underwent a 40 min dynamic ^{18}F-FET PET were retrospectively included, consisting of 22 negative and 62 positive cases (43 low-grade (LGG) and 19 high-grade (HGG) gliomas) as confirmed by histopathology/follow-up. Ground truth labels for lesion and background segmentations were defined by a nuclear medicine physician using MIM software (version 7.0.3). A kinetic filtering approach was considered for a voxel-wise classification into 4 different kinetic classes including blood pool, background tissue (BG), LGG, or HGG. In addition, a UNet-based CNN model was trained to segment tumoral lesions using the 20-40 min static scan by minimizing a weighted combination of cross-entropy and soft Dice loss with the latter adapted to handle the empty labels of negative cases. Both approaches were implemented using three-fold cross-validation with 67% of data used for training and 33% for testing. ***Results:*** KF could not differentiate between negative and positive scans as it detected tumoral voxels in all ^{18}F-FET PET scans. However, KF accurately estimated BG uptake compared to a manual estimation (linear regression (no intercept): slope=0.98, R^2=0.98). On the other hand, the CNN model achieved a sensitivity of 83.64% and precision of 92% (false positive/negative = 4/9 cases) for discriminating between positive and negative ^{18}F-FET PET scans and was able to accurately extract the maximal lesion uptake compared to a manual extraction (linear regression (no intercept): slope=1.0, R^2=1). Combining KF and CNN for extracting BG and maximal lesion uptake respectively resulted in an accurate estimation of MI compared to the fully manual approach (linear regression (no intercept): slope=0.99, R^2=0.98). ***Conclusion:*** The proposed CNN model can detect positive ^{18}F-FET PET scans with high sensitivity and precision. When combined with KF, it also allows for an automatic and accurate MI estimation, therefore significantly reducing the time/workload needed to analyze ^{18}F-FET PET scans.