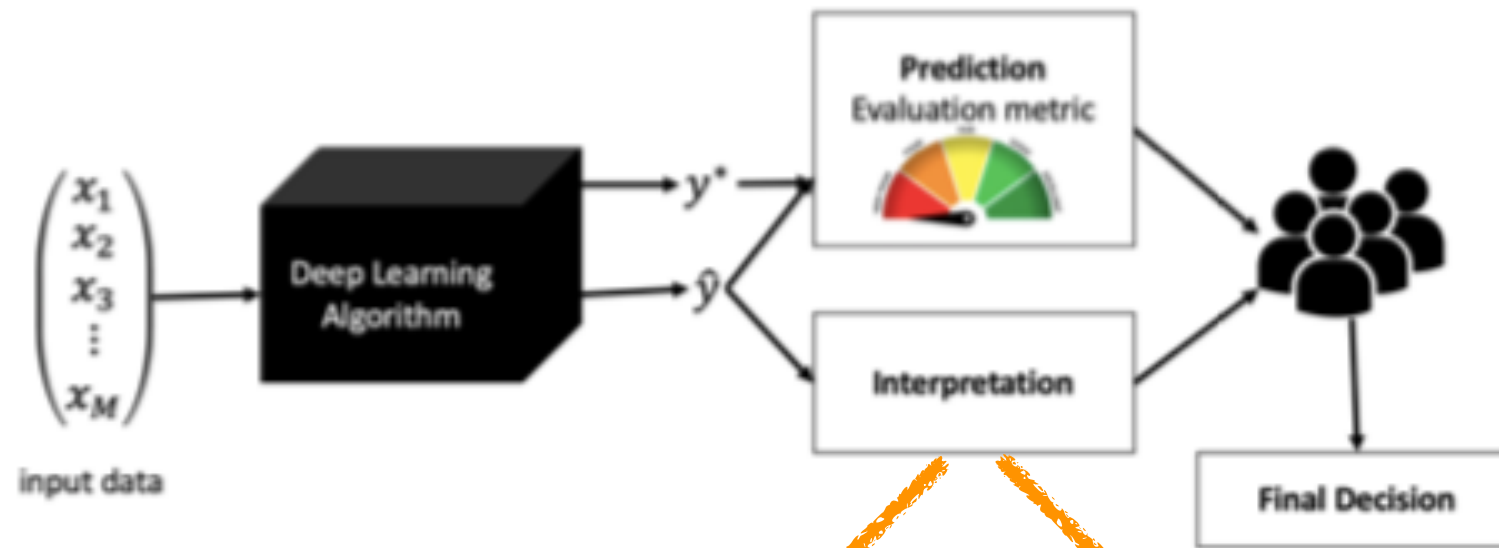# EXPLANATORY ARTIFICIAL INTELLIGENCE (XAI) IN THE PREDICTION OF POST-OPERATIVE LIFE EXPECTANCY IN LUNG CANCER PATIENTS

Pharmacology

Manu Siddhartha	Sr. Manager-IT, Digital Training Department, Allahabad Bank, Head Office, Kolkata, India
Paramita Maity	R.F. Engineer, Tagore Technologies Pvt. Ltd. Kolkata, India
Rajendra Nath	Professor, Department of Pharmacology & Therapeutics, King George's Medical University, Lucknow, India

An Investigation of Interpretability Techniques for Deep Learning in Predictive Process Analytics

Catarina Moreira, Renuka Sindhgatta, Chun Ouyang, Peter Bruza, and Andreas Wichert,
{catarina.pintomoreira, renuka.sr, c.ouyang, p.bruza}@qut.edu.au,
andreas.wichert@tecnico.ulisboa.pt,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_M \end{pmatrix}$$

input data

Deep Learning Algorithm

y^*

$\hat{y}$

Prediction
Evaluation metric

Interpretation

Final Decision

SHAP values

LIME

S.No.	Attribute	Description	Type
1.	PRE 14	T in clinical TNM (size of the original tumor, from OC11 (smallest) to OC14 (largest))	OC11,OC12, OC13 and OC14
2.	DGN	Diagnosis (specific combination of ICD-10 codes for primary and secondary as well multiple tumors if any)	DGN3,DGN2 ,DGN4,DGN 6,DGN5,DG N8,DGN1
3.	PRE4	Forced vital capacity (FVC)	Numeric
4.	PRE5	Volume that has been exhaled at the end of the first second of forced expiration (FEV1)	True, False
5.	PRE6	Performance status – zubrod scale	PRZ0, PRZ1,
6.	PRE7	Pain (pre-surgery)	True, False
7.	PRE8	Haemoptysis (pre-surgery)	True, False
8.	PRE9	Dyspnoea	True, False
9.	PRE10	Cough (pre-surgery)	True, False
10.	PRE11	Weakness (pre-surgery)	True, False
11.	PRE17	Type 2 DM (Diabetes Mellitus)	True, False
12.	PRE19	MI up to 6 months	True, False
13.	PRE25	PAD (Peripheral Arterial Diseases)	True, False
14.	PRE30	Smoking	True, False
15.	PRE32	Asthma	True, False
16.	AGE	Age at surgery in years	Numeric
17.	RISK 1Y	1 year survival period - (T)rue value if died	True, False

Table 4. Accuracy, Recall, F1- Score and Specificity measures of different variants of Random Forest Algorithm

S.No.	Evaluation metrics	Perfect Model	Imbalanced Model
1.	Accuracy (10 Fold Cross validated)	84.04	83.11
2.	Recall / Sensitivity	94.28	81
3.	F1- Score	89.79	76
4.	Specificity	97.25	95
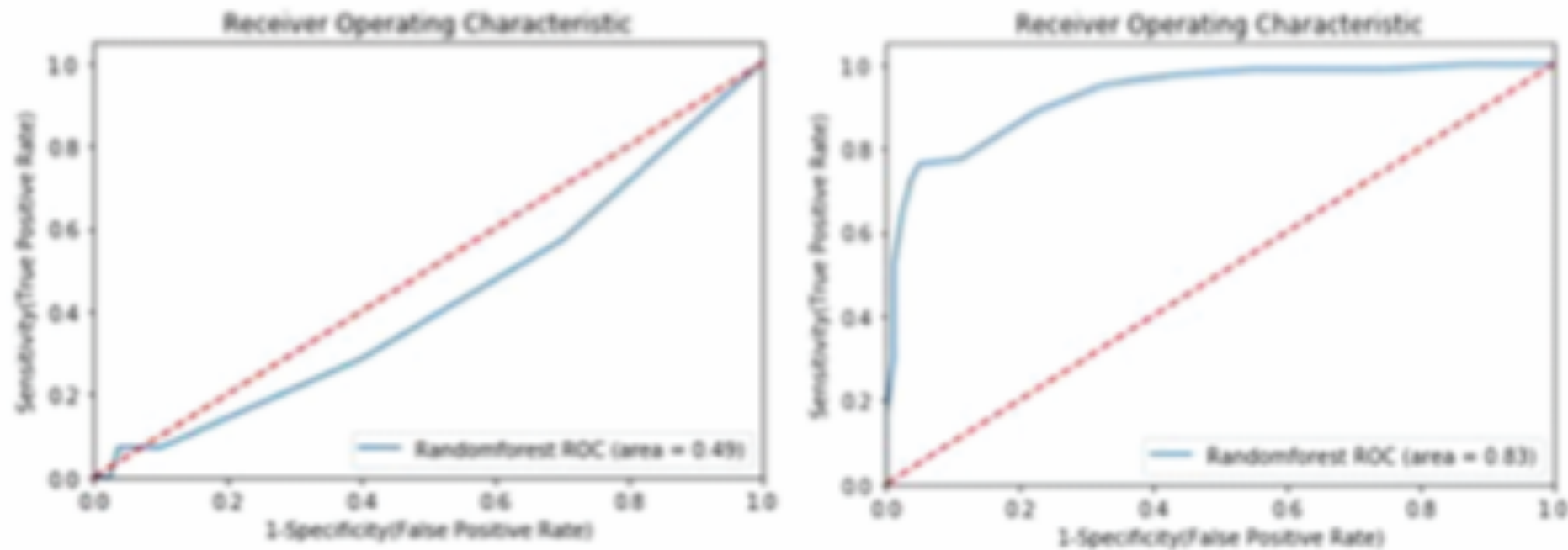
Fig. 5 Receiver Operating Characteristic Curve (ROC) of perfect random forest model (right) with area under curve is 0.83 and for imbalanced random forest model (left) is 0.49
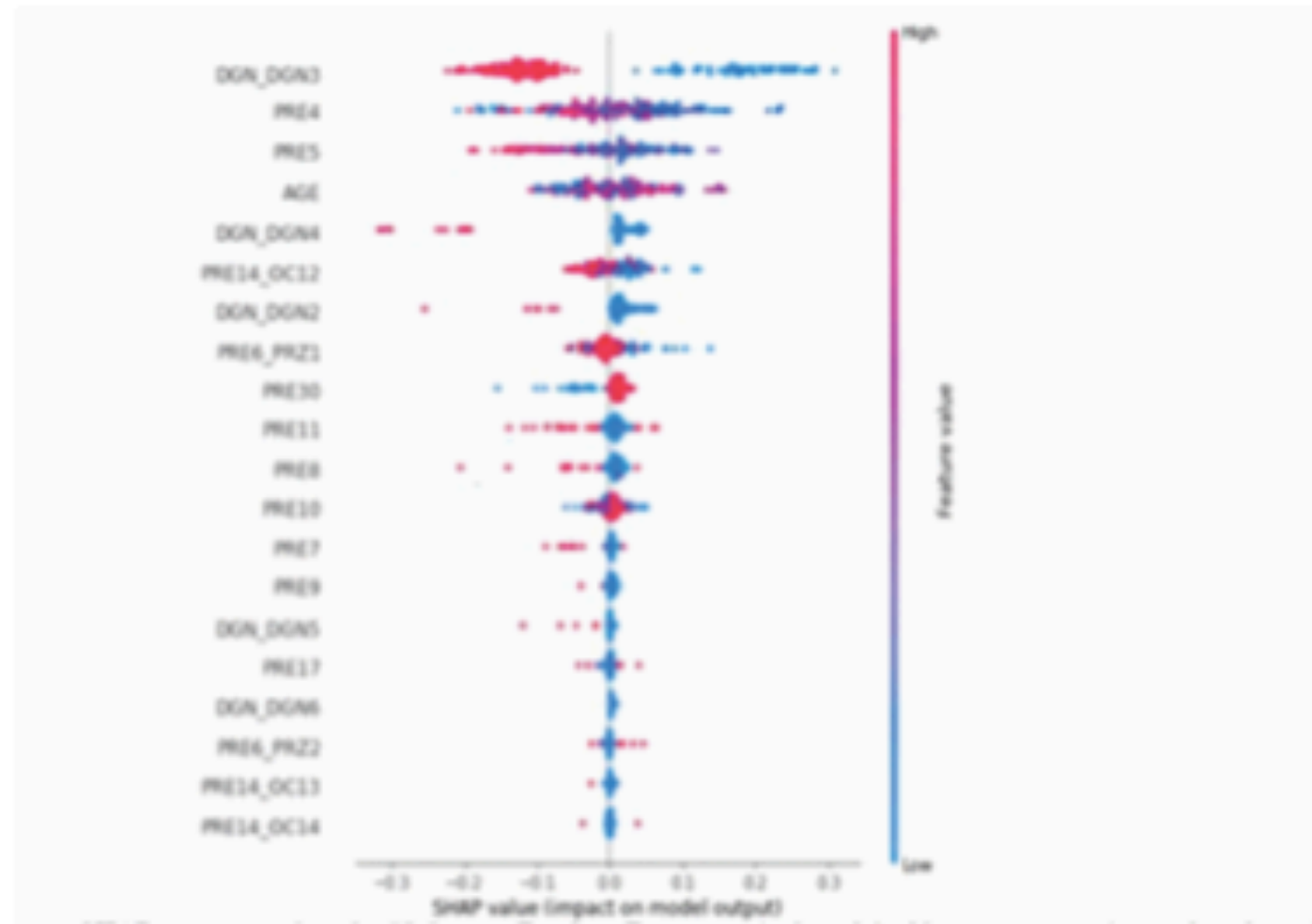
SHAP values

Fig-6 a

Fig-6 b
Fig 6. The SHAP explanation force plot for individual instances

DGN	ICD10
PRE10	Kaszel
Pre4	Spirometria
Pre_14	T - rozmiar guza
Pre30	Palenie
Pre5	FEV1

SHAP summary

Fig 7. SHAP summary plot of a 20 feature Random Forest survival model of lung cancer patients after thoracic surgery. The higher the SHAP value of a feature, the higher is the log odds of death in this lung cancer model. Every patient in the dataset is run through the model and a dot is created for each feature attribution value, so one patient gets one dot on each feature's line. Dot's are colored by the feature's value for that patient and pile up vertically to show density.

LIME - Local Surrogate
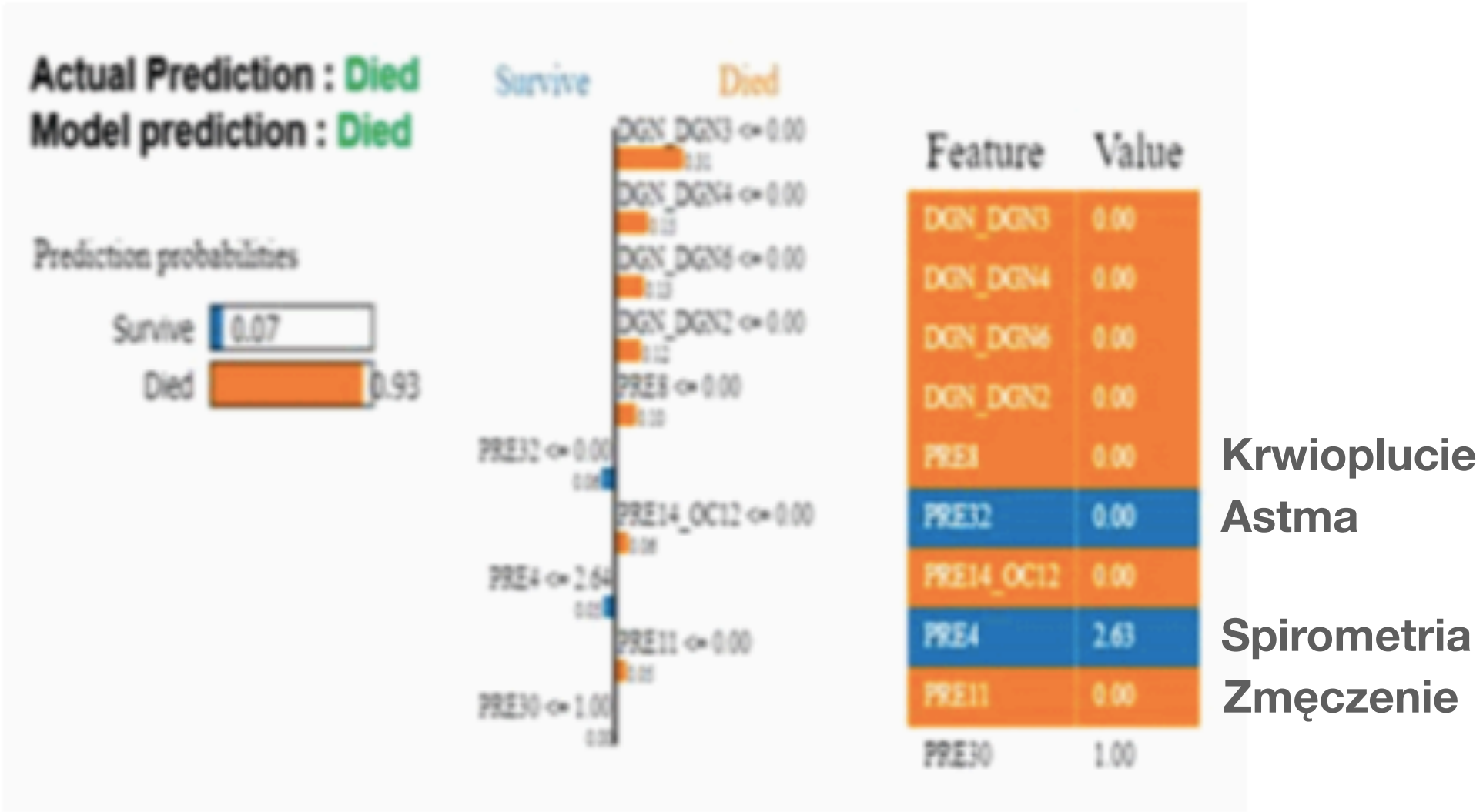
Fig. 8 The local explanation from LIME. The left figure is prediction probabilities of two classes. The middle figure is the LIME explanation of selected features. The representation of numerical features are discretized features. The right table shows the original feature values.

| Patient ID	Activity ID	Timestamp	Cancer ID	Age
1515	A	25-05-2005T16:15	M13	22
1515	B	25-05-2005T16:45	M13	22
1515	A	27-05-2005T09:30	M13	22
1515	C	25-05-2005T16:15	M13	22
...	...	...	...	...

Medical Event Log

$$X^{(1)} = \begin{pmatrix} f_{1,1}^{(1)} & f_{1,2}^{(1)} & \cdots & f_{1,F}^{(1)} \\ f_{2,1}^{(1)} & f_{2,2}^{(1)} & \cdots & f_{2,F}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ f_{T,1}^{(1)} & f_{T,2}^{(1)} & \cdots & f_{1,F}^{(1)} \end{pmatrix}_{T \times F} \cdots$$

$$X^{(M)} = \begin{pmatrix} f_{1,1}^{(M)} & f_{1,2}^{(1)} & \cdots & f_{1,F}^{(M)} \\ f_{2,1}^{(M)} & f_{2,2}^{(1)} & \cdots & f_{2,F}^{(M)} \\ \vdots & & \ddots & \vdots \\ f_{T,1}^{(M)} & f_{T,2}^{(1)} & \cdots & f_{1,F}^{(M)} \end{pmatrix}_{T \times F}$$

$$Y = \begin{pmatrix} class^{(1)} \\ class^{(2)} \\ \vdots \\ class^{(M)} \end{pmatrix}_{M \times 1}$$

Activity	Aktywność, np. pobranie krwi, radioterapia
Department	Oddział związany z aktywnością
Number of executions	Ile razy wykonana była aktywność
Activity code	-
Producer code	-
Section	-
Age	Wiek
Diagnosis Code	Kod przypisany rodzajowi nowotworu
Treatment Code	-
Diagnosis Treatment Combination ID	
Years	
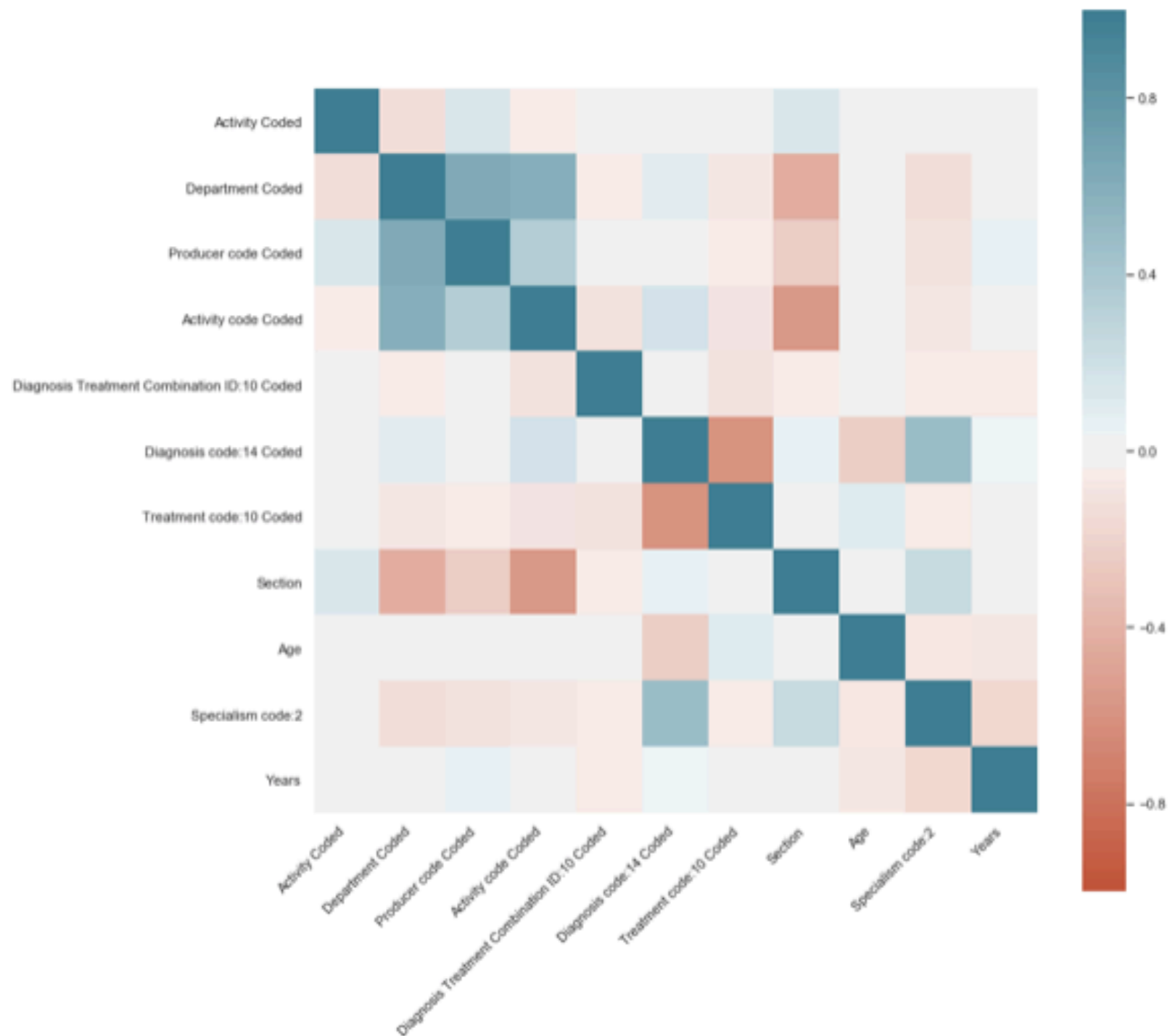
Estimators	Maximum features	Accuracy
1000	100	0.556
1500	200	0.572

TABLE III: Results obtained while conducting a cross validation grid search over the the number of estimators and size of the random subsets of features used for splitting a node in the tree.
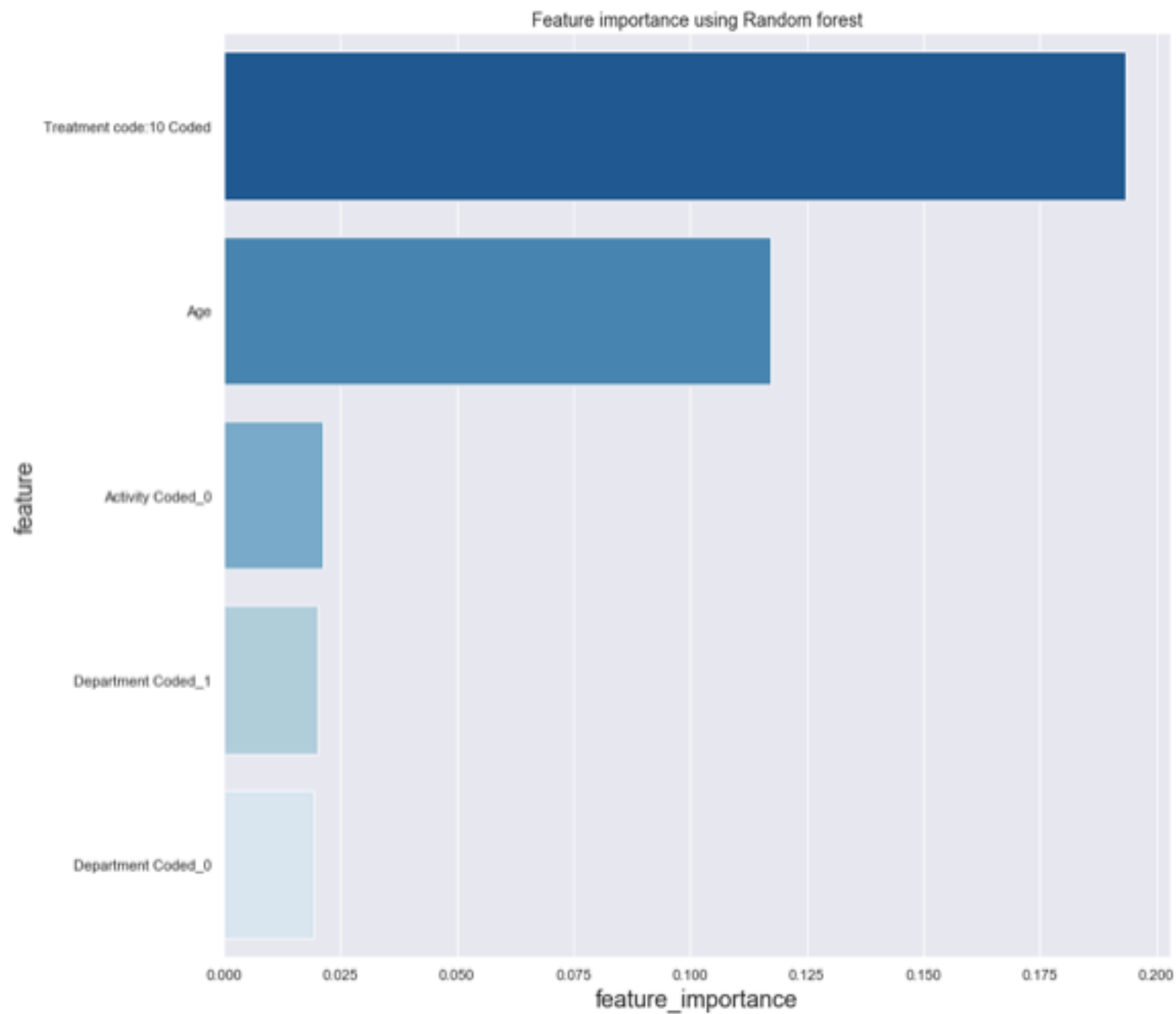
Fig. 11: Top 5 important features used by Random Forest classifier.
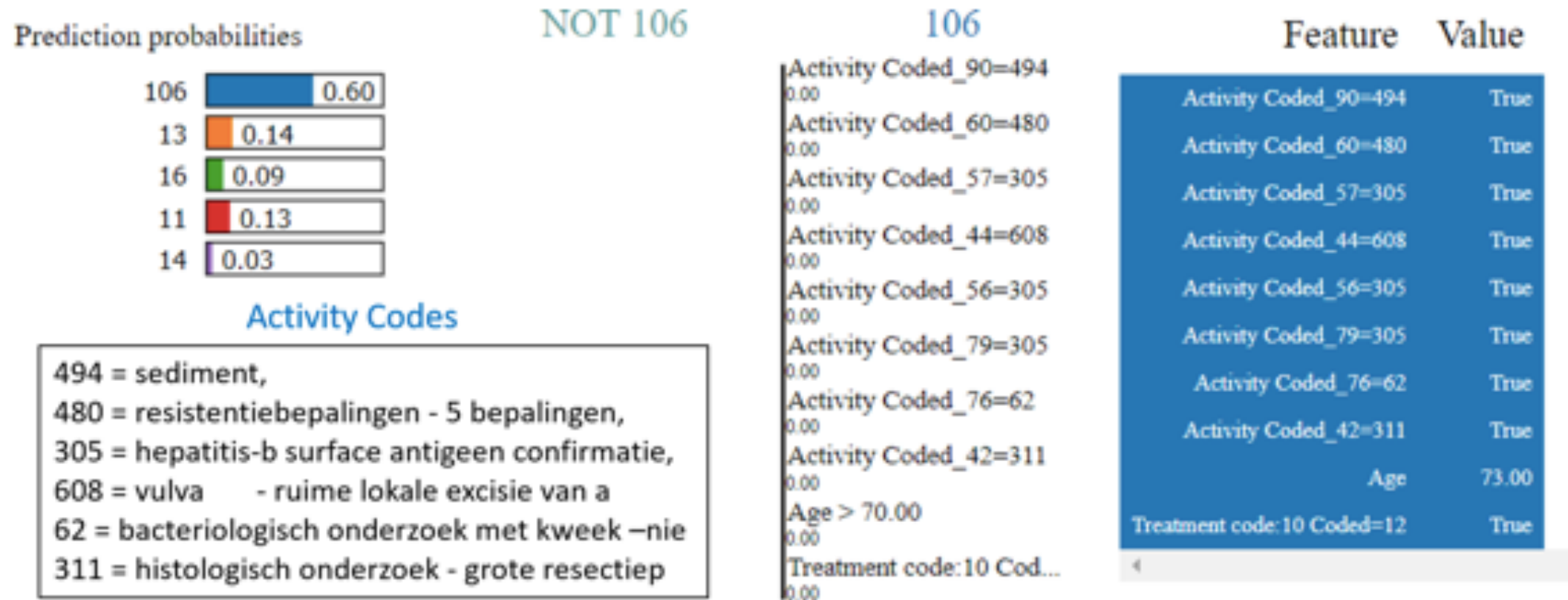
LIME

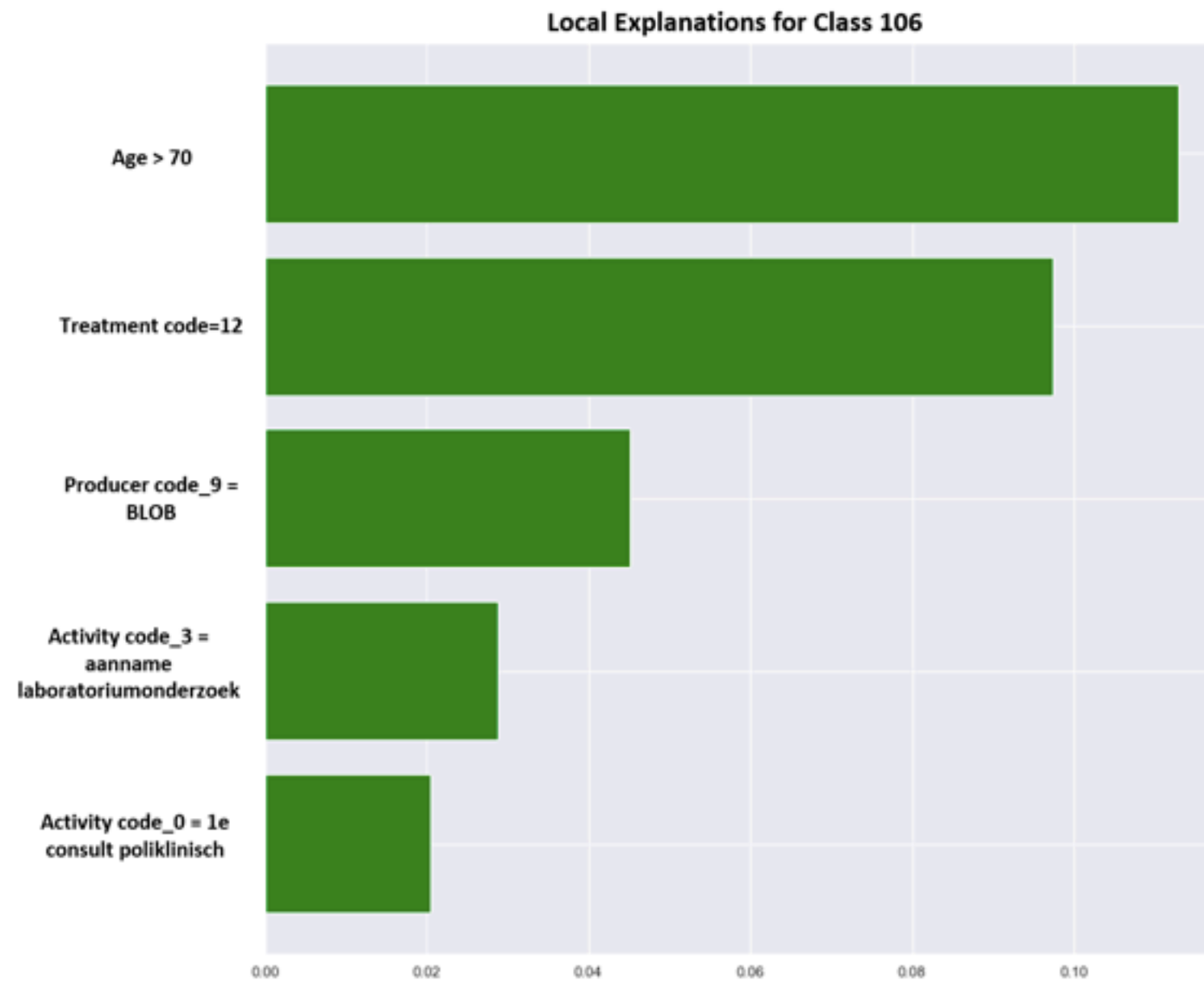Fig. 12: Local interpretation of 106 cancer class for a patient.

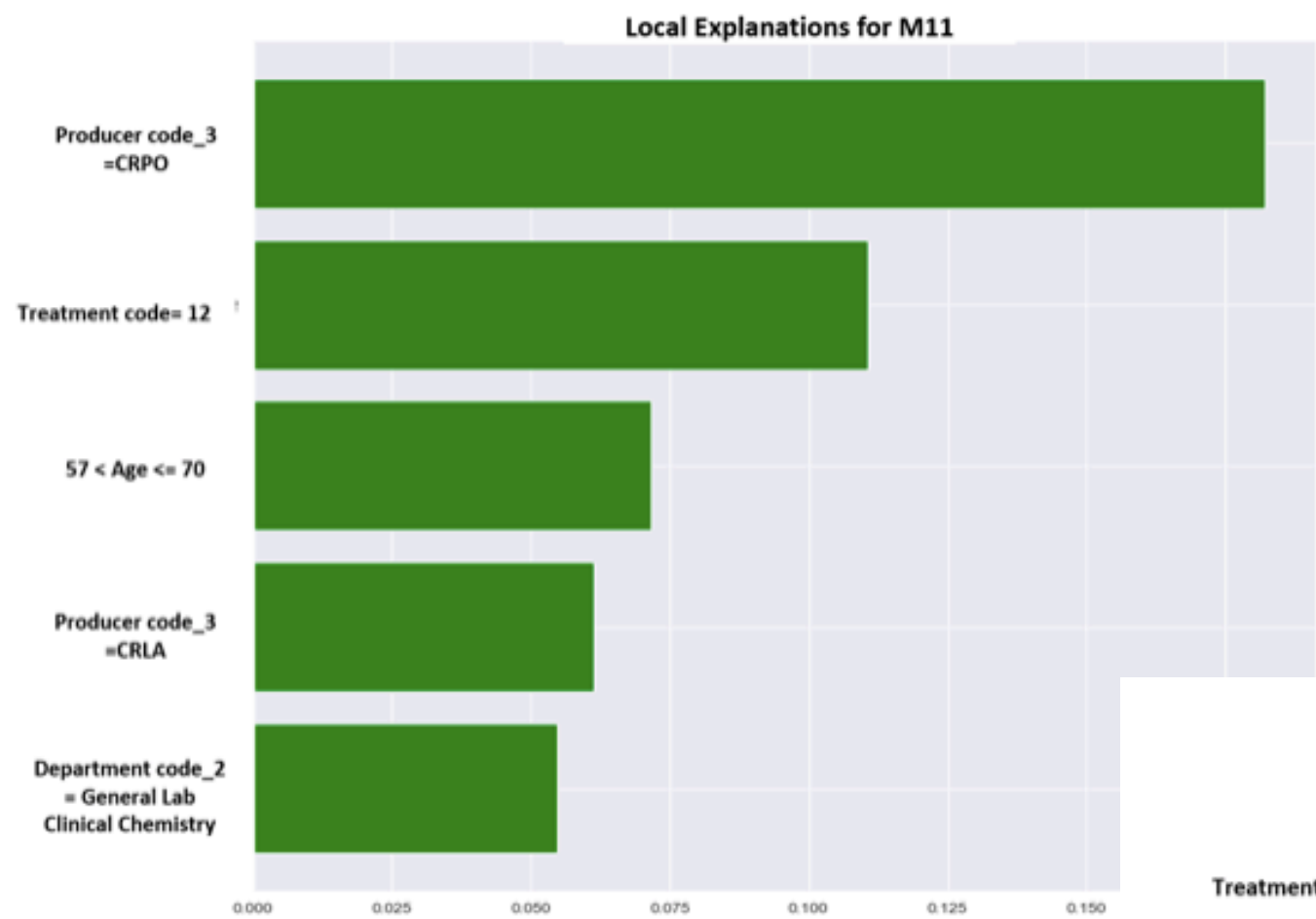Fig. 13: Global interpretation of class 106 cancer.
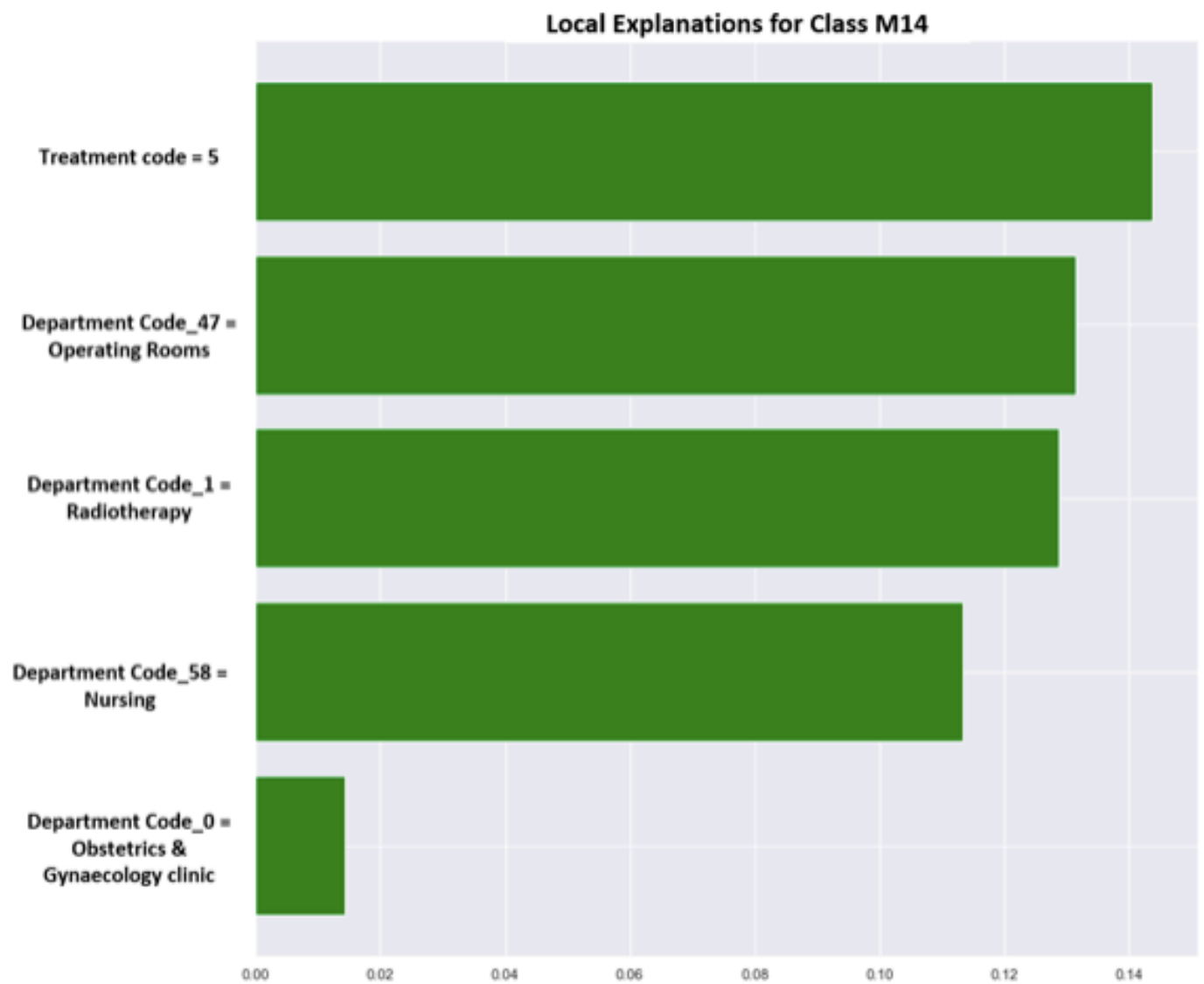
Fig. 14: Global interpretation of class M11 cancer.

Fig. 15: Global interpretation of class M14 cancer.