

Supplementary Material: paper entitled “Computer-aided Diagnosis of Vertebral Compression Fractures using Convolutional Neural Networks and Radiomics”, Journal of Digital Imaging

R. S. Del Lama et al.

January 31, 2022

This is the supplementary material to the paper entitled “Computer-aided Diagnosis of Vertebral Compression Fractures using Convolutional Neural Networks and Radiomics”, submitted to Journal of Digital Imaging. This material provides: i) Genetic Algorithm codification (Section 1); ii) ROC curves from each model (Section 2); iii) The average execution times for each model and the configuration of the machine used in the experiments (Section 3); iv) The learning and architecture parameters of the best model found in each approach (Section 4); v) The features selected by the GA (Section 5).

1 Genetic Algorithm Codification

Figure 1 shows how the hyper-parameters of the CNN and features are codified in the chromosome of the individual.

2 Results

2.1 Model using only radiomics features, and histogram and clinical information

2.1.1 MLP

Table 1 shows the precision results (for the test set) obtained by the best MLP model. The outputs of the MLP for each class, multiplied by 100, are

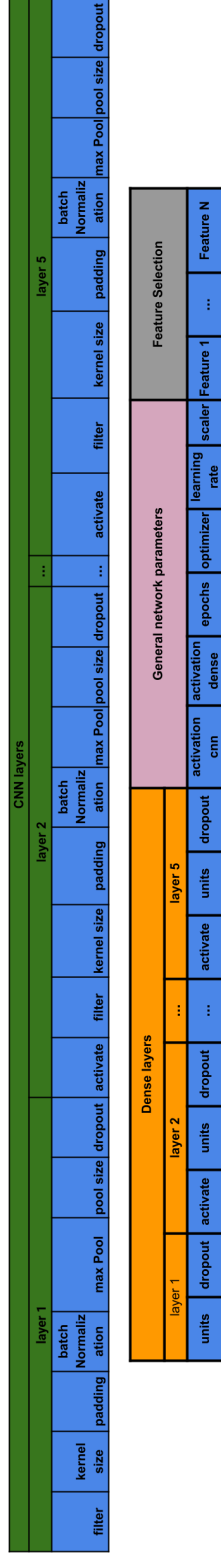


Figure S 1: Representation of the solution (values for the hyper-parameters of the CNN and features) in the chromosome of the individual.

also shown. In the first column, the index after “P” indicates the patient number, and the index after “L” indicates the vertebral body number. In Table 1, the misclassified examples of the test set are shown in boldface.

Table S 1: Results of the MLP for the test set. The MLP obtained an balanced accuracy of 58.8%.

Example	Real Class	Predicted	Benign	Malignant	Normal
P2L1	benign	benign	84.61%	3.35%	12.04%
P2L2	normal	normal	10.31%	1.24%	88.45%
P2L3	benign	normal	41.99%	3.09%	54.92%
P2L4	normal	normal	4.25%	1.78%	93.97%
P2L5	benign	normal	34.37%	4.80%	60.83%
P16L2	malignant	malignant	0.06%	99.90%	0.04%
P16L4	malignant	malignant	0.04%	99.95%	0.01%
P20L1	benign	malignant	18.87%	76.43%	4.7%
P20L2	normal	normal	3.30%	1.45%	95.25%
P20L3	normal	normal	5.62%	1.06%	93.32%
P20L4	normal	normal	2.28%	2.49%	95.23%
P20L5	normal	normal	1.40%	0.42%	98.18%
P24L1	malignant	benign	71.58%	8.63%	19.79%
P24L2	malignant	benign	66.72%	20.43%	12.85%
P33L1	benign	normal	13.40%	2.32%	84.28%
P33L2	normal	normal	0.44%	0.4%	99.16%
P33L3	normal	normal	1.10%	1.78%	97.12%
P33L4	benign	normal	9.98%	2.82%	87.20%
P33L5	normal	normal	1.76%	1.39%	96.85%
P50L3	malignant	malignant	32.77%	62.89%	4.34%

Table S 2: Confusion matrix obtained by the MLP.

	Classification	Predicted Class		
		Benign	Malignant	Normal
Real Class	Benign	1	1	4
	Malignant	2	3	0
	Normal	0	0	9

2.2 Models using only (raw) images

2.2.1 Manually optimized CNN

Tables 3 and 4 and Figure 3 shows the results (for the test set) obtained by manually optimized CNN.

2.2.2 Manually optimized CNN with data augmentation

Tables 5 and 6 and Figure 4 shows the results (for the test set) obtained by manually optimized CNN with data augmentation.

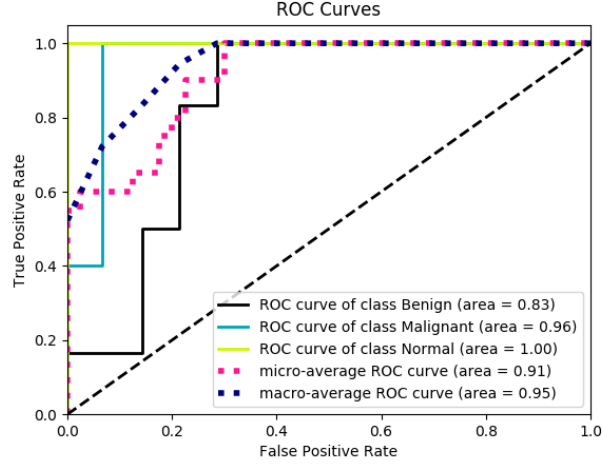


Figure S 2: ROC curves for the MLP.

2.2.3 Pre-Trained CNN

Tables 7 and 8 and Figure 5 shows the results (for the test set) obtained by the pre-trained CNN (VGG16).

2.3 Proposed hybrid models (using all sources of information)

2.3.1 Hybrid model using pre-trained CNN

Tables 9 and 10 and Figure 6 shows the results (for the test set) obtained by the hybrid model using pre-trained CNN.

2.3.2 Hybrid model optimized by the Genetic Algorithm

Tables 11 and 12 and Figure 7 shows the results (for the test set) obtained by the hybrid model optimized by the GA.

Table S 3: Results for the manually optimized CNN. The CNN obtained an balanced accuracy of 77.3%.

Example	Real Class	Predicted	Benign	Malignant	Normal
P2L1	benign	benign	95.88%	4.07%	0.05%
P2L2	normal	normal	0.00%	0.00%	100%
P2L3	benign	malignant	8.17%	52.67%	39.16%
P2L4	normal	normal	39.81%	6.54%	53.65%
P2L5	benign	benign	99.98%	0.02%	0.00%
P16L2	malignant	malignant	0.19%	99.78%	0.01%
P16L4	malignant	malignant	0.1%	99.90%	0.00%
P20L1	benign	malignant	0.16%	99.83%	0.01%
P20L2	normal	normal	0.00%	0.00%	100%
P20L3	normal	normal	0.00%	0.00%	100%
P20L4	normal	normal	0.12%	0.02%	99.86%
P20L5	normal	normal	0.00%	0.00%	100%
P24L1	malignant	malignant	28.80%	70.50%	0.70%
P24L2	malignant	benign	92.20%	7.25%	0.55%
P33L1	benign	benign	100%	0.00%	0.00%
P33L2	normal	normal	0.00%	0.00%	100%
P33L3	normal	normal	0.01%	0.30%	99.59%
P33L4	benign	normal	0.16%	5.39%	94.45%
P33L5	normal	normal	0.00%	0.00%	100%
P50L3	malignant	malignant	1.85%	98.15%	0.00%

Table S 4: Confusion matrix obtained by the manually optimized CNN.

	Classification	Predicted Class		
		Benign	Malignant	Normal
Real Class	Benign	3	2	1
	Malignant	1	4	0
	Normal	0	0	9

3 Typical execution times for the best models

Table 13 shows the average time per run for each model¹. The experiments were carried out on 2 machines: i) a computer with an Intel i5-8400 processor (with 9 MB Cache and 4.00 GHz), 16 GB of RAM and GTX 1060 video card (6 GB of dedicated memory); ii) a server with 2 Intel Xeon E5-2620 v2 processors (with 15 MB Cache and 2.10 GHz) and 32 GB of RAM.

4 Parameters for the Best Models

This section presents the learning and architecture parameters for the best model found in each approach. All models were trained using the *categori-*

¹The ROC Curve, Confusion Matrix, and other results for the test set were generated during the run, which can impact the running time. This explains why the test execution time was longer than the training time in some cases.

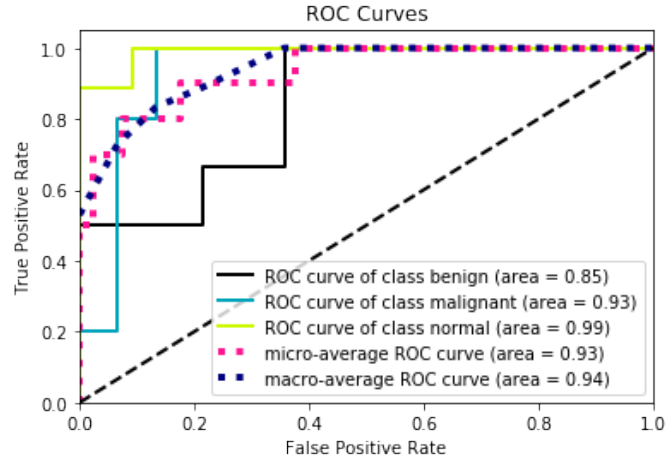


Figure S 3: ROC curves for the manually optimized CNN.

cal_accuracy metric function and the loss *categorical_crossentropy* function. The pre-trained CNNs were trained with RGB images; for this reason the models that used the *VGG16* were trained with 3-channel images. Tables 14 - 18 present the layers and their respective parameters for all models.

5 FEATURES SELECTED BY THE GENETIC ALGORITHM

Tables 19 and 20 shows the features selected by the best solution found in each of the 5 runs of the Genetic Algorithm.

Table S 5: Results for the manually optimized CNN with data augmentation. The CNN obtained an balanced accuracy of 75.5%.

Example	Real Class	Predicted	Benign	Malignant	Normal
P2L1	benign	benign	76.56%	23.40%	0.04%
P2L2	normal	normal	0.20%	0.87%	98.93%
P2L3	benign	malignant	19.07%	80.55%	0.38%
P2L4	normal	normal	20.28%	2.34%	77.38%
P2L5	benign	benign	94.31%	5.69%	0.00%
P16L2	malignant	malignant	22.46%	77.24%	0.03%
P16L4	malignant	malignant	0.01%	99.39%	0.6%
P20L1	benign	benign	52.41%	99.83%	0.01%
P20L2	normal	normal	4.22%	5.38%	90.4%
P20L3	normal	normal	5.04%	0.00%	94.96%
P20L4	normal	normal	1.15%	0.00%	98.85%
P20L5	normal	normal	0.05%	0.11%	99.84%
P24L1	malignant	malignant	41.98%	51.69%	6.33%
P24L2	malignant	normal	1.01%	1.33%	97.66%
P33L1	benign	benign	81.87%	17.14%	0.99%
P33L2	normal	normal	0.21%	0.19%	99.6%
P33L3	normal	normal	9.13%	0.00%	90.87%
P33L4	benign	normal	3.76%	6.35%	89.89%
P33L5	normal	normal	0.27%	6.74%	92.99%
P50L3	malignant	benign	57.73%	41.86%	0.41%

Table S 6: Confusion matrix obtained by the manually optimized CNN with data augmentation.

	Classification	Predicted Class		
		Benign	Malignant	Normal
Real Class	Benign	4	1	1
	Malignant	1	3	1
	Normal	0	0	9

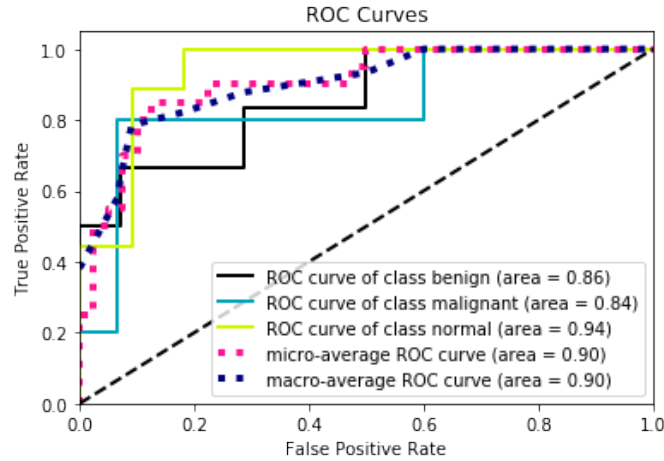


Figure S 4: ROC curves for the manually optimized CNN with data augmentation.

Table S 7: Results for the pre-trained CNN (VGG16). The CNN obtained an balanced accuracy of 82.96%.

Example	Real Class	Predicted	Benign	Malignant	Normal
P2L1	benign	benign	99.66%	0.06%	0.28%
P2L2	normal	normal	0.02%	0.01%	99.97%
P2L3	benign	benign	91.83%	2.84%	5.33%
P2L4	normal	normal	0.25%	0.05%	99.70%
P2L5	benign	benign	75.94%	0.26%	23.80%
P16L2	malignant	benign	43.95%	43.80%	12.25%
P16L4	malignant	malignant	8.00%	87.80%	4.20%
P20L1	benign	benign	57.03%	35.74%	7.23%
P20L2	normal	normal	0.23%	3.10%	96.67%
P20L3	normal	normal	0.20%	0.18%	99.62%
P20L4	normal	normal	1.40%	2.62%	95.98%
P20L5	normal	normal	0.01%	0.01%	99.98%
P24L1	malignant	benign	77.16%	29.74%	3.10%
P24L2	malignant	malignant	35.51%	58.56%	5.93%
P33L1	benign	benign	95.25%	3.17%	1.58%
P33L2	normal	normal	0.14%	0.50%	99.36%
P33L3	normal	malignant	8.56%	71.26%	20.18%
P33L4	benign	benign	73.99%	14.34%	11.67%
P33L5	normal	normal	0.04%	0.12%	99.84%
P50L3	malignant	malignant	9.29%	86.72%	3.99%

Table S 8: Confusion matrix obtained by the pre-trained CNN (VGG16).

	Classification	Predicted Class		
		Benign	Malignant	Normal
Real Class	Benign	6	0	0
	Malignant	2	3	0
	Normal	0	1	8

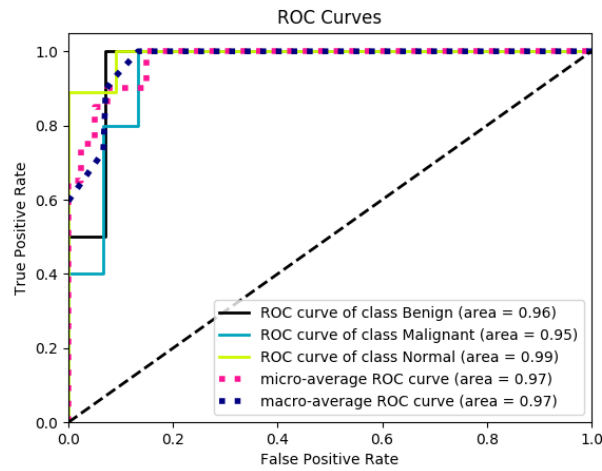


Figure S 5: ROC curves for the pre-trained CNN (VGG16).

Table S 9: Results for the hybrid model using pre-trained CNN. The hybrid model obtained an balanced accuracy of 87.77%.

Example	Real Class	Predicted	Benign	Malignant	Normal
P2L1	benign	benign	96.60%	3.22%	0.19%
P2L2	normal	normal	0.30%	0.05%	99.65%
P2L3	benign	benign	74.01%	24.16%	1.83%
P2L4	normal	normal	0.08%	0.01%	99.91%
P2L5	benign	benign	78.83%	0.76%	20.41%
P16L2	malignant	malignant	0.96%	98.35%	0.69%
P16L4	malignant	malignant	0.36%	99.52%	0.12%
P20L1	benign	malignant	14.83%	84.90%	0.27%
P20L2	normal	normal	0.12%	0.64%	99.24%
P20L3	normal	normal	0.15%	0.16%	99.69%
P20L4	normal	normal	0.06%	0.1%	99.84%
P20L5	normal	normal	0.00%	0.00%	1.00%
P24L1	malignant	benign	51.11%	48.35%	0.54%
P24L2	malignant	malignant	39.80%	59.06%	1.14%
P33L1	benign	benign	90.52%	6.74%	2.74%
P33L2	normal	normal	0.05%	0.08%	99.87%
P33L3	normal	normal	26.70%	16.50%	56.80%
P33L4	benign	benign	76.69%	4.72%	18.59%
P33L5	normal	normal	0.01%	0.01%	99.98%
P50L3	malignant	malignant	33.19%	66.57%	0.24%

Table S 10: Confusion matrix obtained by the hybrid model using pre-trained CNN.

	Classification	Predicted Class		
		Benign	Malignant	Normal
Real Class	Benign	5	1	0
	Malignant	1	4	0
	Normal	0	0	9

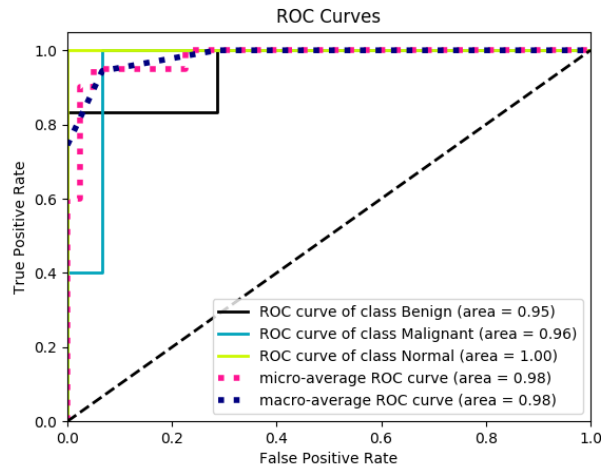


Figure S 6: ROC curves for the hybrid model using pre-trained CNN.

Table S 11: Results for the hybrid model optimized by the GA. The hybrid model obtained an balanced accuracy of 70%.

Example	Real Class	Predicted	Benign	Malignant	Normal
P2L1	benign	benign	99.34%	0.25%	0.41%
P2L2	normal	normal	0.00%	0.01%	99.99%
P2L3	benign	normal	30.22%	1.14%	68.64%
P2L4	normal	normal	0.16%	0.02%	99.82%
P2L5	benign	normal	00.80%	0.26%	98.94%
P16L2	malignant	malignant	0.04%	99.94%	0.02%
P16L4	malignant	malignant	0.01%	99.98%	0.01%
P20L1	benign	benign	56.18%	42.85%	0.97%
P20L2	normal	normal	0.43%	0.10%	99.47%
P20L3	normal	normal	15.20%	0.02%	84.78%
P20L4	normal	normal	4.68%	0.01%	95.31%
P20L5	normal	normal	0.73%	0.01%	99.26%
P24L1	malignant	benign	99.55%	00.42%	0.03%
P24L2	malignant	benign	96.60%	3.03%	0.37%
P33L1	benign	benign	78.13%	0.10%	21.77%
P33L2	normal	normal	0.41%	0.00%	99.59%
P33L3	normal	normal	4.86%	0.02%	95.12%
P33L4	benign	normal	8.22%	0.06%	91.72%
P33L5	normal	normal	0.25%	0.01%	99.74%
P50L3	malignant	malignant	22.09%	77.87%	0.04%

Table S 12: Confusion matrix obtained by the hybrid model optimized by the GA.

	Classification	Predicted Class		
		Benign	Malignant	Normal
Real Class	Benign	3	0	3
	Malignant	2	3	0
	Normal	0	0	9

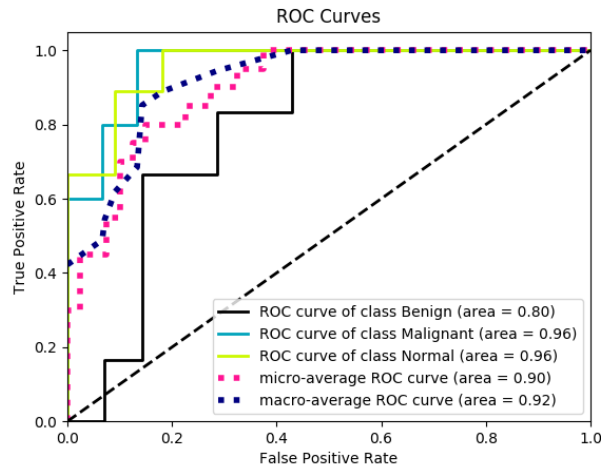


Figure S 7: ROC curves for the hybrid model optimized by the GA.

Table S 13: Average time per run for each of the models used in this work.

Model	Cross-validation	Test	GA	Hardware used
MLP	1 min	6 min	-	Hardware ii
CNN manually optimized	6 min	8 min	-	Hardware i
CNN manually optimized with data augmentation	13 min	13 min	-	Hardware i
Pre-trained CNN	50 min	106 min	-	Hardware ii
Hybrid model with pre-trained CNN	53 min	275 min	-	Hardware ii
Hybrid model optimized by the GA	3 min	4 min	85 h	Hardware ii

Table S 14: Parameters for the MLP. The model was trained for 100 epochs, using the Adam optimizer with learning rate (lr) 0.001 and the data was normalized (standard scaling).

Layer	Parameters
Dense	units=113, input_dim=113
Dense	units=58
Dropout	rate=0.2
Dense	units=29
Dropout	rate=0.2
Dense	units=3

Table S 15: Parameters for the CNN manually optimized. The model was trained for 50 epochs, using the Adam optimizer with $lr = 0.001$.

Layer	Parameters
Conv2D	filters=128, kernel_size=3x3, input_shape=(125, 91, 1)
BatchNormalization	-
MaxPooling2D	pool_size=2x2
Conv2D	filters=64, kernel_size=3x3
BatchNormalization	-
MaxPooling2D	pool_size=2x2
Conv2D	filters=64, kernel_size=3x3
BatchNormalization	-
MaxPooling2D	pool_size=2x2
Flatten	-
Dense	units=256
Dropout	rate=0.4
Dense	units=64
Dropout	rate=0.4
Dense	units=3

Table S 16: Parameters of the pre-trained CNN (VGG16). The model was re-trained for 100 epochs, using the Adam optimizer with $lr = 0.0001$.

Layer	Parameters
VGG16	input_shape=(100, 100, 3)
Flatten	-
Dense	units=512
Dropout	rate=0.2
Dense	units=256
Dropout	rate=0.2
Dense	units=128
Dropout	rate=0.2
Dense	units=3

Table S 17: Parameters of the hybrid model with pre-trained CNN (VGG16). The model was pre-trained for 300 epochs, using the Adam optimizer with $lr = 0.0001$ and the external features data was normalized (standard scaling).

Layer	Parameters
VGG16	input_shape=(100, 100, 3)
Flatten	-
Concatenate Flatten + 113 Features adicionais	-
Dense	units=628
Dropout	rate=0.2
Dense	units=314
Dropout	rate=0.2
Dense	units=157
Dropout	rate=0.2
Dense	units=3

Table S 18: Parameters of the hybrid model optimized by the GA. The model was trained for 150 epochs, using the SGD optimizer with $lr = 0.05$ and the the external features data was normalized (standard scaling). The learning and architecture parameters were found by the GA.

Layer	Parameters
Conv2D	filters=128, kernel_size=4x4, padding='valid', input_shape=(150, 150, 1), activation='tanh'
BatchNormalization	-
MaxPooling2D	pool_size=5x5
Dropout	rate=0.3
Conv2D	filters=512, kernel_size=3x3, padding='same', activation='tanh'
Dropout	rate=0.1
Conv2D	filters=64, kernel_size=6x6, padding='valid', activation='tanh'
BatchNormalization	-
MaxPooling2D	pool_size=3x3
Dropout	rate=0.1
Flatten	-
Concatenate Flatten + 53 Features adicionais	-
Dense	units=200
Dropout	rate=0.2
Dense	units=200
Dropout	rate=0.3
Dense	units=100
Dropout	rate=0.1
Dense	units=3

Table S 19: Features selected by the best solution found in each of the 5 runs of the GA. The total of features is 113.

Feature	Run 1	Run 2	Run 3	Run 4	Run 5
age	X	X	X	X	X
sex			X		
Max	X	X	X	X	
Mean	X		X	X	
Min		X		X	X
STD			X	X	X
vertebral body	X		X	X	
firstorder_10Percentile	X		X		X
firstorder_90Percentile	X		X		X
firstorder_Energy	X	X	X	X	
firstorder_Entropy		X	X		X
firstorder_InterquartileRange	X	X	X	X	
firstorder_Kurtosis			X	X	
firstorder_Maximum			X	X	
firstorder_Mean		X	X	X	
firstorder_Median				X	X
firstorder_MeanAbsoluteDeviation	X		X		
firstorder_Minimum	X	X	X		
firstorder_Range	X		X	X	
firstorder_RobustMeanAbsoluteDeviation	X	X	X	X	X
firstorder_RootMeanSquared			X		
firstorder_StandardDeviation		X	X		X
firstorder_Skewness	X	X	X		X
firstorder_TotalEnergy	X			X	X
firstorder_Uniformity	X	X			
firstorder_Variance	X		X		
glcm_Autocorrelation		X	X	X	
glcm_ClusterProminence				X	
glcm_ClusterTendency					X
glcm_ClusterShade	X	X		X	
glcm_Contrast	X	X			X
glcm_Correlation	X	X		X	X
glcm_DifferenceAverage			X	X	
glcm_DifferenceEntropy	X	X		X	
glcm_DifferenceVariance	X	X		X	
glcm_Id			X		
glcm_Idm		X	X		X
glcm_Idn	X	X			
glcm_Idmn			X		
glcm_Imc1		X	X		X
glcm_Imc2	X				
glcm_InverseVariance		X	X		X
glcm_JointAverage					X
glcm_JointEnergy	X	X	X	X	
glcm_JointEntropy					X
glcm_MaximumProbability	X		X		
glcm_SumAverage	X			X	
glcm_SumEntropy		X			X
glcm_SumSquares		X	X		X
gldm_DependenceEntropy		X	X		
gldm_DependenceNonUniformity		X			X
gldm_DependenceNonUniformityNormalized				X	X
gldm_DependenceVariance	X	X		X	X
gldm_GrayLevelNonUniformity	X		X		X
gldm_GrayLevelVariance		X	X		X
gldm_HighGrayLevelEmphasis	X	X			
gldm_LargeDependenceEmphasis		X		X	
gldm_LargeDependenceHighGrayLevelEmphasis	X	X			
gldm_LargeDependenceLowGrayLevelEmphasis	X				
gldm_LowGrayLevelEmphasis	X				X
gldm_SmallDependenceEmphasis		X		X	
gldm_SmallDependenceHighGrayLevelEmphasis	X			X	X

Table S 20: (Continuation) Features selected by the best solution found in each of the 5 runs of the GA. The total of features is 113.

Feature	Run 1	Run 2	Run 3	Run 4	Run 5
glrlm_GrayLevelNonUniformity			X	X	
glrlm_GrayLevelNonUniformityNormalized				X	X
glrlm_GrayLevelVariance	X				X
glrlm_HighGrayLevelRunEmphasis	X	X		X	
glrlm_LongRunEmphasis	X	X		X	
glrlm_LongRunHighGrayLevelEmphasis	X	X		X	
glrlm_LowGrayLevelRunEmphasis	X				
glrlm_LongRunLowGrayLevelEmphasis		X		X	
glrlm_RunEntropy	X		X	X	X
glrlm_RunLengthNonUniformity	X				X
glrlm_RunLengthNonUniformityNormalized	X	X		X	
glrlm_RunPercentage X			X		
glrlm_RunVariance	X				
glrlm_ShortRunEmphasis	X	X		X	X
glrlm_ShortRunHighGrayLevelEmphasis	X	X	X		
glrlm_ShortRunLowGrayLevelEmphasis		X	X		
glszm_GrayLevelNonUniformity				X	X
glszm_GrayLevelNonUniformityNormalized					X
glszm_GrayLevelVariance				X	X
glszm_HighGrayLevelZoneEmphasis	X	X	X		X
glszm_LargeAreaEmphasis	X	X		X	X
glszm_LargeAreaHighGrayLevelEmphasis		X	X		
glszm_LargeAreaLowGrayLevelEmphasis	X	X	X		X
glszm_LowGrayLevelZoneEmphasis		X			X
glszm_SizeZoneNonUniformity	X	X			X
glszm_SizeZoneNonUniformityNormalized	X			X	
glszm_SmallAreaHighGrayLevelEmphasis	X			X	X
glszm_SmallAreaLowGrayLevelEmphasis	X			X	
glszm_ZoneEntropy	X		X	X	X
glszm_ZonePercentage				X	X
glszm_ZoneVariance	X			X	X
ngtdm_Busyness			X		X
ngtdm_Coarseness	X		X	X	
ngtdm_Contrast				X	
ngtdm_Strength	X				X
shape_ElongationX					
shape_Flatness X			X		X
shape_LeastAxis	X			X	
shape_MajorAxis		X	X		X
shape_Maximum2DDiameterColumn	X	X		X	
shape_Maximum2DDiameterRow	X	X		X	X
shape_Maximum2DDiameterSlice		X			X
shape_Maximum3DDiameter	X		X		X
shape_MinorAxis	X		X		
shape_Compactness1		X	X	X	X
shape_Compactness2			X		X
shape_Sphericity	X			X	X
shape_SphericalDisproportion	X				X
shape_SurfaceArea		X			
shape_SurfaceVolumeRatio	X				
shape_Volume		X		X	
Total selected features	65	56	50	55	56