

Confidential until publication in Scientific Report

Prediction Models of Tumor Growth Trajectories Based on Pretreatment CT Images of TKI-treated Lung Cancer Patients

Authors:

Huy Gia Truong¹, Hidetaka Arimura^{2*}, Kentaro Tanaka^{3*}, Kenta Ninomiya¹, Quoc Cuong Le¹

Affiliations:

¹ Division of Medical Quantum Science, Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University

² Division of Medical Quantum Science, Department of Health Sciences, Faculty of Medical Sciences, Kyushu University

³ Respiratory Medicine, Kyushu University Hospital

Corresponding authors:

Hidetaka Arimura, PhD, Division of Medical Quantum Science, Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

[E-mail: arimurah@med.kyushu-u.ac.jp]

Kentaro Tanaka, MD, PhD, Department of Respiratory Medicine, Kyushu University Hospital, 3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

[E-mail: tanaka-k@kokyu.med.kyushu-u.ac.jp]

Table S1. Spearman correlations between the number of tumor cells derived from CT images and the number of tumor cells with mean absolute percentage errors (MAPEs) for an SVR model in a training dataset.

TKI	Patient ID	Spearman correlation	p-value	MAPE (%)
Erlotinib	Genelky079 (1 st treatment)	0.800	0.200	6.64
	Genelky023	0.678	0.094	55.6
	Genelky021 (3 rd treatment)	0.800	0.200	4.39
	Genelky034	0.429	0.337	83.1
	Average	0.677	0.208	37.4
Afatinib	Genelky117 (1 st treatment)	0.900	0.037	14.7
	Genelky039 (1 st treatment GTV1)	1.000	0	16.6
	Genelky041	0.964	4.54×10^{-4}	8.30
	Genelky007	0.750	0.020	18.3
	Genelky060	0.976	3.31×10^{-5}	7.66
	Genelky124 (GTV1)	0.483	0.187	18.6
	Genelky124 (GTV2)	0.950	8.76×10^{-5}	5.10
	Average	0.861	0.035	12.7
Gefitinib	Genelky068 (1 st treatment)	0.679	0.094	3.10
	Genelky045 (1 st treatment GTV1)	0.929	0.003	7.06
	Genelky021 (1 st treatment)	1.000	0	1.94
	Genelky117 (2 nd treatment)	1.000	1.40×10^{-24}	1.87
	Genelky088	0.857	0.014	4.67
	Genelky093	0.967	2.16×10^{-5}	11.5
	Genelky091 (GTV2)	0.771	0.072	6.51
	Genelky091 (GTV1)	0.829	0.042	8.52
	Average	0.879	0.028	5.65
Osimertinib	Genelky127	0.571	0.139	8.82
	Genelky045 (2 nd treatment GTV1)	0.750	0.052	12.3
	Genelky045 (2 nd treatment GTV2)	0.893	0.007	8.13
	Genelky002	0.900	0.037	3.51
	Genelky071	0.976	3.31×10^{-5}	14.0
	Genelky101 (3 rd treatment)	0.900	0.037	8.17
	Genelky056	0.500	0.391	20.3
	Genelky079 (3 rd treatment)	0.933	2.36×10^{-4}	9.00
	Average	0.803	0.083	10.5
Overall average		0.822	0.073	13.6

Table S2. Spearman correlations between the number of tumor cells derived from CT images and the number of tumor cells with mean absolute percentage errors (MAPEs) for an RF model in training dataset.

TKI	Patient ID	Spearman correlation	p-value	MAPE (%)
Erlotinib	Genelky079 (1 st treatment)	0.800	0.200	15.0
	Genelky023	-0.714	0.071	1.77×10^3
	Genelky021 (3 rd treatment)	0.400	0.600	12.7
	Genelky034	0.429	0.337	63.4
	Average	0.229	0.302	466
Afatinib	Genelky117 (1 st treatment)	0.700	0.188	18.3
	Genelky039 (1 st treatment GTV1)	1.000	0	42.8
	Genelky041	0.857	0.014	21.1
	Genelky007	0.433	0.244	52.0
	Genelky060	0.976	3.31×10^{-5}	23.7
	Genelky124 (GTV1)	0.250	0.516	45.7
	Genelky124 (GTV2)	0.867	0.002	6.52
	Average	0.726	0.138	30.0
Gefitinib	Genelky068 (1 st treatment)	0.143	0.760	57.2
	Genelky045 (1 st treatment GTV1)	0.929	0.003	63.5
	Genelky021 (1 st treatment)	0.400	0.600	27.4
	Genelky117 (2 nd treatment)	0.900	0.037	38.5
	Genelky088	0.750	0.052	26.6
	Genelky093	0.950	8.76×10^{-5}	8.51
	Genelky091 (GTV2)	0.771	0.072	9.54
	Genelky091 (GTV1)	0.486	0.329	68.0
	Average	0.666	0.232	37.4
Osimertinib	Genelky127	0.214	0.610	56.3
	Genelky045 (2 nd treatment GTV1)	0.536	0.215	47.9
	Genelky045 (2 nd treatment GTV2)	0.500	0.253	50.7
	Genelky002	0.800	0.104	7.65
	Genelky071	0.714	0.047	32.0
	Genelky101 (3 rd treatment)	0.100	0.873	27.6
	Genelky056	0.800	0.104	18.4
	Genelky079 (3 rd treatment)	0.833	0.005	16.8
	Average	0.562	0.276	32.2
Overall average		0.586	0.231	97.4

Table S3. Spearman correlations between the number of tumor cells derived from CT images and the number of tumor cells with mean absolute percentage errors (MAPEs) for an SVR model in a test dataset.

TKI	Patient ID	Spearman correlation	p-value	MAPE (%)
Erlotinib	Genelky067	0.821	0.023	29.2
	Genelky101 (2 nd treatment)	0.762	0.028	17.0
	Average	0.792	0.026	23.1
Afatinib	Genelky046	0.786	0.021	68.8
	Genelky016	0.905	0.002	8.12
	Average	0.845	0.011	38.5
Gefitinib	Genelky072 (GTV2)	1.000	1.40×10^{-24}	8.28
	Genelky072 (GTV3)	0.800	0.104	73.4
	Genelky072 (GTV1)	0.900	0.037	15.0
	Average	0.900	0.047	32.2
Osimertinib	Genelky039 (2 nd treatment GTV2)	0.905	0.002	55.4
	Genelky068 (2 nd Treatment)	0.607	0.148	24.1
	Genelky039 (2 nd treatment GTV1)	0.690	0.058	32.8
	Average	0.734	0.069	37.5
Overall average		0.818	0.042	33.2

Table S4. Spearman correlations between the number of tumor cells derived from CT images and the number of tumor cells with mean absolute percentage errors (MAPEs) for an RF model in a test dataset.

TKI	Patient ID	Spearman correlation	p-value	MAPE (%)
Erlotinib	Genelky067	0.393	0.383	44.2
	Genelky101 (2 nd treatment)	0.857	0.007	17.4
	Average	0.625	0.195	30.8
Afatinib	Genelky046	0.619	0.102	67.8
	Genelky016	0.810	0.015	13.4
	Average	0.714	0.058	40.6
Gefitinib	Genelky072 (GTV2)	0.900	0.037	7.48
	Genelky072 (GTV3)	0.300	0.624	171
	Genelky072 (GTV1)	0.700	0.188	8.26
	Average	0.633	0.283	62.4
Osimertinib	Genelky039 (2 nd treatment GTV2)	0.595	0.120	68.7
	Genelky068 (2 nd Treatment)	0.607	0.148	22.5
	Genelky039 (2 nd treatment GTV1)	0.738	0.037	14.1
	Average	0.647	0.101	35.1
Overall average		0.652	0.166	43.5

Table S5. 54 image feature types calculated from pre-treatment CT images.

Histogram-based features	Texture features			
	Gray level co-occurrence matrix (GLCM) ¹	Gray level run-length matrix (GLRLM) ²	Gray level size-zone matrix (GLSZM) ³	Neighborhood gray-tone difference matrix (NGTDM) ⁴
<ul style="list-style-type: none"> • Energy • Entropy • Kurtosis • Maximum • Mean • Mean absolute difference • Median • Minimum • Range • Root mean square • Skewness • Standard deviation (SD) • Uniformity • Variance 	<ul style="list-style-type: none"> • Auto correlation • Contrast • Correlation • Dissimilarity • Energy • Entropy • Homogeneity • Variance • Sum average 	<ul style="list-style-type: none"> • Gray level non-uniformity (GLN) • Gray level variance (GLV) • High gray level run emphasis (HGRE) • Long run emphasis (LRE) • Long run high gray level emphasis (LRHGE) • Long run low gray level emphasis (LRLGE) • Low gray level run emphasis (LGRE) • Run length non-uniformity (RLN) • Run-length variance (RLV) • Run percentage (RP) • Short run emphasis (SRE) • Short run high gray level emphasis (SRHGE) • Short run low gray level emphasis (SRLGE) 	<ul style="list-style-type: none"> • Gray level non-uniformity (GLN) • Gray level variance (GLV) • High gray level zone emphasis (HGZE) • Large zone emphasis (LZE) • Large zone high gray level emphasis (LZHGE) • Large zone low gray level emphasis (LZLGE) • Low gray level zone emphasis (LGZE) • Small zone emphasis (SZE) • Small zone high gray level emphasis (SZHGE) • Small zone low gray level emphasis (SZLGE) • Zone percentage (ZP) • Zone-size non-uniformity (ZSN) • Zone-size variance (ZSV) 	<ul style="list-style-type: none"> • Busyness • Coarseness • Complexity • Contrast • Strength
n=14	n=9	n=13	n=13	n=5
			n=40	
		n=54		

Table S6. Interpretation range for Spearman correlation coefficient⁵

Size of Correlation	Interpretation
0.90 to 1.00 (-0.90 to -1.00)	Very high positive (negative) correlation
0.70 to 0.90 (-0.70 to -0.90)	High positive (negative) correlation
0.50 to 0.70 (-0.50 to -0.70)	Moderate positive (negative) correlation
0.30 to 0.50 (-0.30 to -0.50)	Low positive (negative) correlation
0.00 to 0.30 (0.00 to -0.30)	Negligible correlation

References

1. Haralick, R. M., Shanmugam, K., & Dinstein, I. H. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, **6**, pp: 610-621 (1973).
2. Galloway, M. M. Texture analysis using gray level run lengths. *STIN*, **75**, 18555 (1975).
3. Thibault, G. Texture indexes and gray level size zone matrix. Application to cell nuclei classification. *In Proc. Pattern Recognition and Information Processing*. Minsk, Belarus, pp: 140–145 (2009).
4. Amadasun, M., King, R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern.* **19**, pp:1264–1274. doi:10.1109/21.44046 (1989).
5. Mukaka, M. M. Statistics corner: a guide to appropriate use of correlation in medical research. *Malawi Medical Journal*, **24**(3), pp. 69-71 (2012).