

厚生労働科学研究費補助金(臨床研究等ICT基盤構築・人工知能実装研究事業)

総括 研究報告書

ビッグデータからの機械学習による前立腺癌小線源療法の予後予測法の開発と均てん化への応用

研究代表者 中村 和正 浜松医科大学 医学部 教授

研究要旨：

小線源療法にて治療されたJ-POPSコホート1, 2の前立腺癌6430例において、経過観察期間の中央値5.00年(0.07-10.8年)にて、原病死13例、他病死167例であった。NCCNリスク分類での5年PSA非再発率は、低リスク群98.3%、中リスク群95.2%、高リスク群90.8%で、リスクが上がるほど有意にPSA非再発率が低かった。これらの症例の予後予測精度を向上するために、ロジスティック回帰分析にて有意であったGleasonスコア、T因子、PSA値、Total BEDの4つの因子にて、回帰係数より各々の症例でのPSA非再発確率を求め、PSA非再発確率が0.978860以上をRisk A、0.889500未満をRisk C、中間をRisk Bとして、新たなリスク分類とした。各々のリスク群の5年PSA非再発率は、Risk A 99.3%、Risk B 95.8%、Risk C 83.5%とNCCNリスク分類よりも明確にPSA再発を予測できた。さらに、NCCNリスク分類の各リスク群を本計算式で分類することにより、各リスク群の中からよりPSA再発のリスクの高い群を抽出することが可能となった。

一方、ロジスティック回帰、SVM、RF、DNNの機械学習手法を用いて、PSA再発の予測を改善できるかを検討した。いずれも学習データでは予測精度は0.65 - 0.9以上と高い値を示したが、テスト症例の予測精度は現時点ではいずれも低かった。特に、Precision、すなわち再発と予測した症例のうち、真に再発した症例の割合はいずれも0.15以下であった。

今回の研究で明らかとなった計算式により、PSA非再発率をより正確に予測できることが判明した。

研究分担者

斉藤史郎 東京医療センター・医長
萬 篤憲 東京医療センター・医長
片山敬久 岡山大学・放射線科
馬込大貴 駒沢大学・講師
小島伸介 医療イノベーション推進センター・TRI専門職
菊池 隆 医療イノベーション推進センター・上席研究員

A . 研究目的

前立腺癌の放射線治療の予後因子とし

て、T因子、PSA値、Gleason分類などがある。これらによって低・中・高リスクに分類され、大まかな予後予測が可能で、それに沿ったノモグラムが作成されている。しかし、それ以外の患者背景や検査所見、照射線量、治療パラメータなどの多くの因子を網羅的に分析し、治療結果を予測する方法については国内外を含めてほとんど研究が進められていない。

我々は、ヨウ素125シード線源を用いた小線源療法に関する前向きコホート研究（JPOPS, Japanese Prostate Cancer Outcome Study of Permanent I-125 seed Implantation）を実施してきた。本研究には全国74施設（小線源療法を施行する施設の約70%）が参加し、2005年から2010年までに小線源療法で治療された約7000例（同時期に本邦で小線源療法により治療された症例の約40%）が前向き登録された。調査項目は、患者・家族背景、生活歴、T因子、PSA値、Gleason分類（GS）などの腫瘍因子、小線源治療パラメータ、外照射併用の有無等の治療因子、PSA再発の有無、生死、有害事象の有無等などで、少なくとも5年以上の経過観察が行われている世界的に類を見ないビッグデータである。2005年から2007年までのコホート1の2339例のデータクリーニングが終了し、2010年までのコホート2の約4600例についても平成28年11月で5年の経過観察が終了し、これらのビッグデータが使用可能となった。

本研究の目的は、JPOPSによって得られたビッグデータを用いて、詳細な臨床情報を機械学習させることにより、新しい前立腺癌の予後予測システムを開発し、放射線治療の質の均てん化に資することである。

B．研究方法

JPOPS研究のコホート1（2005-2007年）で収集された約2000例のデータのクリーニングはすでに終了している。また、J-POPS コホート 2 の約 4600 例（2008-2010年）のデータの仮クリーニン

グも令和元年に終了した。平成29年度にはコホート1の患者背景、臨床情報、治療パラメータと、PSA非再発率、全生存率、有害事象発生率等の治療結果との関係について機械学習による評価を開始した。平成30年度には、放射線治療施設の規模と治療の質、JPOPSコホート1での予後因子解析、機械学習による予後予測精度向上の検証、小線源治療後のPSA kineticsの解析を行った。SVM、RF、DNNの機械学習手法にて、PSA再発の予測精度を検証したが、予測精度は0.7-0.9以上と高い値を示したものの、テスト症例の予測精度は現時点ではいずれも低く、機械学習で正確に再発するかどうかを予測することは困難であることが明らかとなった。そのため、T因子、PSA値、GSに加え、単変量解析や過去の論文からPSA非再発率に影響を与えると考えられる、年齢、前立腺体積、Total BED（biological equivalent dose）、外照射有無の計7つの因子について、ロジスティック回帰分析を行い、回帰係数より、各々の症例でのPSA非再発確率 $p(x)$ を求め、PSA非再発確率から、Risk A、Risk B、Risk Cに分類することにより、従来のリスク分類より、明確にPSA再発を予測できることを明らかとした。

これらの研究成果をもとに、令和元年度は本研究を継続し、以下のように研究を進めた。

1) J-POPS コホート 1, 2 全症例での治療成績の解析

J-POPS コホート 2 の約 4600 例（2008-2010年）のデータの仮クリーニングが終了し、J-POPS コホート 1, 2 を合わせた約7000例にて治療成績を解析した。

2) 新しい基準での PSA 非再発率と予後因子との関係

予後予測因子のエンドポイントである PSA 非再発生存率については、国際的に用いられている Phoenix 基準を用いた。PSA 非再発率と予後因子との関係を評価するために、平成 30 年度に実施した、T 因子、PSA 値、GS、年齢、前立腺体積、Total BED、外照射有無の計 7 つの因子について、ロジスティック回帰分析を行った。「結果」に示すように、Gleason 分類、T 因子、PSA 値、BED のみの 4 つの因子が有意となったため、この 4 つの因子を用いて、回帰係数より、各々の症例での PSA 非再発確率 $p(x)$ を求め、PSA 非再発確率 $p(x)$ により、リスク分類を行い、PSA 非再発率を求めた。

3) JPOPS コホート 1、2 での機械学習による予後予測精度向上の検証

POPS コホート 1、2 の約 7000 例のデータを用いて、ロジスティック回帰、サポートベクターマシン(SVM)、ランダムフォレスト(RF)、ディープラーニング(DNN)の機械学習手法を用いて、PSA 再発、有害事象発生率の予測を改善できるかを検討した。入力特徴量は、2) に示した 7 つの因子、4 つの因子に厳選した場合と、となるべく多くの特徴量を採用した Large Database(44 項目)の 2 つのデータセットを用いた。評価法は 10-fold cross-validation 法を用いて Accuracy (全症例にて正しく予測できた症例の割合)、Precision (再発と予測した症例にて真に再発した症例の割合)、Recall (再発症例

にて再発と予測した症例の割合)を評価指標とした。

(倫理面への配慮)

本研究はすでにJPOPSで登録され、匿名化された既存データのみを用いる観察研究であり、患者への侵襲は伴わない。また、JPOPS研究のコホート1およびコホート2のデータセットについては臨床研究情報センターにおいて厳重に管理されている。

本年度、JPOPS研究のコホート1,2のデータを用いて解析を行ったが、データの移送においては、フォルダにパスワードにて暗号化した。また、解析用のパソコンにおいてもパスワード等で厳重に管理されている。

研究の透明性確保のため、研究代表者の所属機関(浜松医科大学)において倫理審査を受け、平成29年7月26日に倫理審査委員会の承認を得ている(研究番号17-095)。また、解析を担当する駒沢大学においても同様に平成29年7月21日に倫理審査委員会の承認を得ている(通知番号17-18)。

匿名化された既存データのみを用いる観察研究ではあるが、今後も、ハード、ソフトウェアおよびデータの移送のすべてのレベルでセキュリティを強固にする。

C . 研究結果

1) J-POPS コホート 1, 2 全症例での治療成績の解析

J-POPS コホート 1, 2 全 6687 例のうち、治療前 PSA、T 因子、Gleason 分類、D90 が入力されており、誤入力等を除いた 6430 例を対象とした。年齢の中央値

69.0 歳 (37-89 歳) 経過観察期間の中央値 5.00 年 (0.07-10.8 年) であった。

全 6430 例にて、原病死 13 例、他病死 167 例であった。PSA 再発については、Phoenix の定義を用い、PSA 再発有りとされた症例は 241 例であった。Phoenix 再発の定義を満たすが、ホルモン無しに低下した例 (218 例) は Bounce とみなし、他病死 167 例とともに、PSA 再発のイベントには含めなかった。

NCCN リスク分類にて、低リスク 2873 例 (44.7%)、中リスク 3067 例 (47.7%)、高リスク(超高リスク含む) 490 例 (7.6%) であった。プレプラン時の前立腺体積の中央値 25.48 cc (6.8 - 179.5 cc) であった。

外照射の有無での投与線量(biologically effective dose, BED)の分布を図 1 に示す。

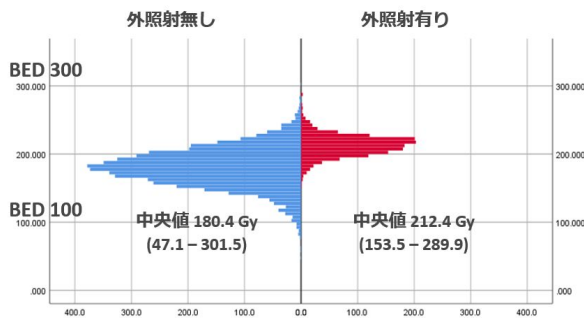


図 1 外照射の有無での BED の分布

各リスク群別の PSA 非再発率を図 2 に示す。

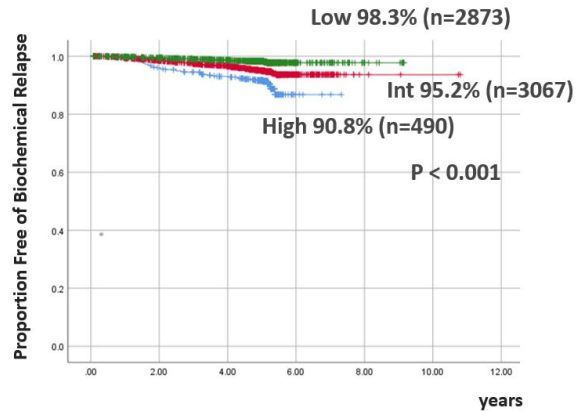


図 2 各リスク群別の PSA 非再発率

5 年 PSA 非再発率は、低リスク群 98.3%、中リスク群 95.2%、高リスク群 90.8%で、リスクが上がるほど有意に PSA 非再発率が低かった ($P < 0.001$)。

2) 新しい基準での PSA 非再発率と予後因子との関係

前年度にコホート 1 の 2316 例で行った解析と同様に、コホート 1, 2 の 6430 例にて、PSA 再発に影響すると考えられる、年齢、Gleason 分類、T 因子、PSA 値、前立腺体積、外照射有無、Total BED の 7 つの因子について、ロジスティック回帰分析を行った。

PSA再発有り=0, PSA再発無し=1として

要因	回帰係数	標準誤差	有意確率
Age	0.007	0.010	0.516
Gleason分類	-0.636	0.095	<0.001
T因子	-0.291	0.065	<0.001
PSA値	-0.040	0.009	<0.001
前立腺体積	-0.011	0.007	0.149
外照射有無	0.319	0.198	0.107
BED	0.011	0.003	<0.001
定数	6.735	1.805	<0.001

- ・ T因子：*TX=1, T1 a=2, T1 b=3, T1 c=4, T2a=5, T2b=6, T2c=7, T3a=8, T3b=9, T4=10
- ・ 外照射有無：無し=1

表 1 7 つの因子でのロジスティック回帰分析

表 1 に示すように、有意な因子は、Gleason 分類、T 因子、PSA 値、BED のみとなった。

そこで、有意差のある、Gleason スコア、T 因子、PSA 値、Total BED の 4 つの因子のみに絞って、ロジスティック回帰分析を行った (表 2)。

PSA再発有り=0, PSA再発無し=1として

要因	回帰係数	標準誤差	有意確率
Gleasonスコア	-0.670	0.088	<0.001
T因子	-0.297	0.064	<0.001
PSA値	-0.041	0.009	<0.001
BED	0.008	0.002	<0.001
定数	7.879	0.616	<0.001

• T因子 : *TX=1, T1 a=2, T1 b=3, T1 c=4, T2a=5, T2b=6, T2c=7, T3a=8, T3b=9, T4=10

表 2 4 つの因子でのロジスティック回帰分析

回帰係数より、各々の症例での PSA 非再発確率 $p(x)$ を下記式より求めた。

$$\log \frac{p(x)}{1-p(x)} = -0.670 \times GS - 0.297 \times T - 0.041 \times PSA + 0.008 \times BED + 7.879$$

$$Z = \log \frac{p(x)}{1-p(x)} \text{ とすると、} p(x) = \frac{1}{1+\exp(-Z)}$$

PSA 非再発確率 $p(x)$ により、0.978860 以上を Risk A、0.889500 未満を Risk C、中間を Risk B として、新たなリスク分類とし、リスク群別の PSA 非再発率曲線を求めた。

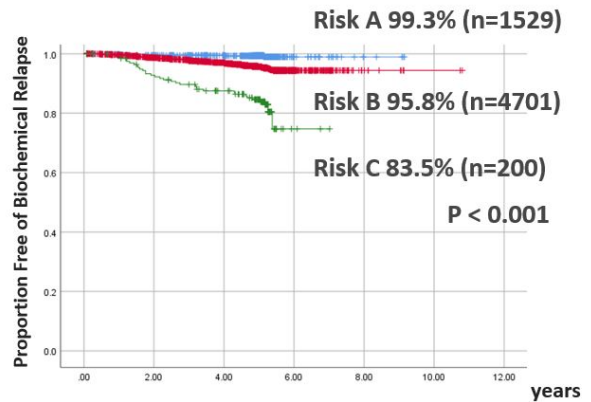


図 3 各リスク群別の PSA 非再発率

5 年 PSA 非再発率は、Risk A 99.3%、Risk B 95.8%、Risk C 83.5%と NCCN リスク分類よりも明確に PSA 再発を予測できた。

さらに、NCCN リスク分類の各リスク群を本計算式で分類した。

低リスク 2873 例では、5 年非再発率は、Risk A (n=1425) 99.4%、Risk B (n=1448) 97.1%、Risk C 該当無しであった (図 4)。

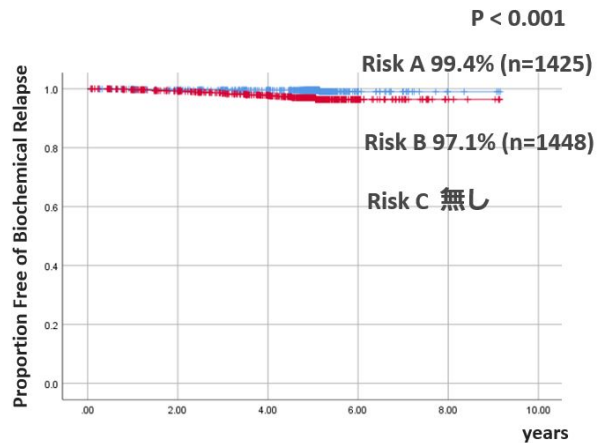


図 4 NCCN 低リスク群での PSA 非再発率

中リスク 3067 例では、5 年非再発率は、Risk A (n=104) 97.1%、Risk B (n=2944) 95.2%、Risk C (n=19) 89.5%

と、有意差はなかったが、中リスク群内をさらにリスク群に分けることが可能であった（図5）。

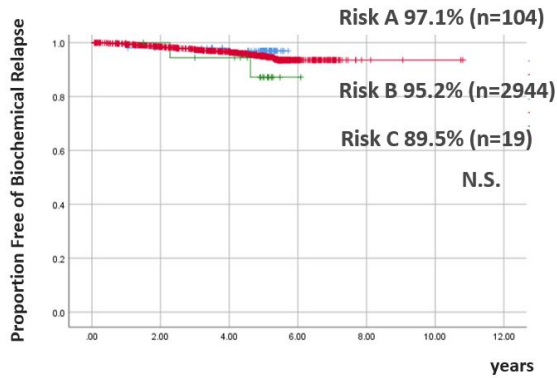


図5 NCCN 中リスク群での PSA 非再発率

高リスク 490 例では、5 年非再発率は、Risk A 該当無し、Risk B(n=309) 95.5%、Risk C (n=181) 82.9%で、有 NCCN リスク高リスク群の中で、より PSA 再発のリスクの高い群を抽出することが可能であった。

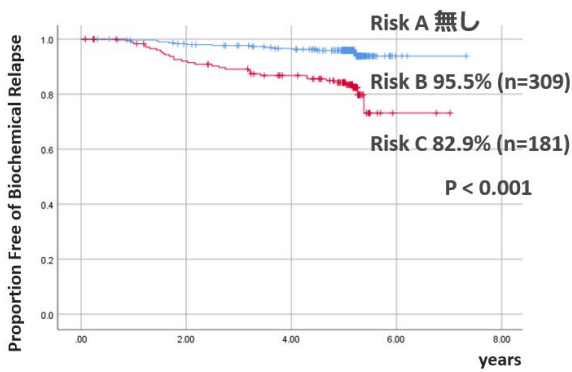


図6 NCCN 高リスク群での PSA 非再発率

3) JPOPS コホート 1、2 での機械学習による予後予測精度向上の検証

Phoenix 基準を用いて、ロジスティック回帰、SVM、RF、DNN の機械学習手法

を用いて、PSA 再発の予測を改善できるかを検討した（表 3）。いずれも学習データでは予測精度は 0.65 - 0.9 以上と高い値を示したが、テスト症例の予測精度は現時点ではいずれも低かった。特に、Precision、すなわち再発と予測した症例のうち、真に再発した症例の割合はいずれも 0.15 以下であった。

		Accuracy		Precision		Recall	
		training	test	training	test	training	test
4因子	Logist	0.642	0.689	0.657	0.069	0.591	0.585
	SVM	0.646	0.680	0.659	0.068	0.606	0.590
	RF	0.990	0.822	0.993	0.059	0.988	0.249
	DNN	0.675	0.550	0.638	0.061	0.810	0.766
7因子	Logist	0.656	0.681	0.665	0.070	0.628	0.604
	SVM	0.646	0.645	0.648	0.064	0.642	0.616
	RF	0.998	0.917	0.999	0.100	0.997	0.148
	DNN	0.681	0.542	0.639	0.060	0.831	0.757
LargeDB	Logist	0.695	0.719	0.704	0.082	0.672	0.633
	SVM	0.666	0.565	0.637	0.061	0.776	0.725
	RF	0.999	0.941	0.999	0.132	0.998	0.083
	DNN	0.766	0.695	0.738	0.075	0.826	0.626

表3 各機械学習での予後予測精度

D. 考察

今回、JPOPSコホート1, 2の約7000例のビッグデータがはじめて解析可能となった。5年PSA非再発率は、NCCNリスク分類にて、低リスク群98.3%、中リスク群95.2%、高リスク群90.8%で、非常に良好な成績であった。そのような中、どのような症例が再発リスクが高いかを正確に予測するアルゴリズムを確立することは極めて重要である。

昨年度、PSA非再発確率pを初歩的な「機械学習」のひとつであるロジスティック回帰分析にて求め、リスク分類することにより、従来のリスク分類より、明確にPSA再発を予測できる可能性があることを明らかとした。

今回、JPOPS コホート 1, 2 のビッグデータを使い、同様の解析を行い、下記の計算式を用いることにより、明確に PSA 再発を予測できることが明らかとなった。

$$\log \frac{p(x)}{1-p(x)} = -0.670 \times GS - 0.297 \times T - 0.041 \times PSA + 0.008 \times BED + 7.879$$

$$Z = \log \frac{p(x)}{1-p(x)} \text{ とすると、 } p(x) = \frac{1}{1+\exp(-z)}$$

本研究の重要性は、従来から、予後因子と言われていた GS, PSA 値、T 因子、投与線量 BED が、それぞれの重み付けをもって、PSA 非再発率に影響していることを明らかとしたことである。

特に、図 6 に示されるように、NCCN 分類の高リスク群を、明確に PSA 再発の確率の高い群と、低い群に分けることができることは意義が大きい。

また、JPOPS コホート 1, 2 のビッグデータを用いて、昨年度同様、SVM、RF、DNN の機械学習手法にて、PSA 再発が予測可能かを検討した。いずれも学習データでは予測精度は 0.7-0.9 以上と高い値を示したものの、テスト症例の予測精度は現時点ではいずれも低く、特に、Precision、すなわち再発と予測した症例のうち、真に再発した症例の割合はいずれも 0.15 以下であった。これは、「再発する」と機械学習で予測しても、15%以下の精度しかないことを意味する。前立腺癌小線源療法においては、JPOPS で取得されたデータのみでは、機械学習を用いても正確に再発するかどうかを予測することは困難であることが判明した。

その原因として、前立腺癌小線源療法では、PSA再発の頻度が極めて少ないため、

予測が難しいこと、および、今回のJPOPSで収集したデータ以外の因子、例えば腫瘍の遺伝子情報、患者の生活習慣等が再発に係わっている可能性があること、の2つを考えている。

E . 結論

今回の研究で明らかとなった計算式により、PSA非再発率をより正確に予測できることが判明した。

以下のような活用が期待できる。

1) 小線源療法にて治療された患者で、治療時点の背景因子、治療パラメータ等を入力することで、従来のリスク分類より、さらに正確にPSA非再発率を予測できる。

2) 治療結果が正確に予測されれば、再発リスクの低い症例では経過観察期間を延ばすことが可能となる。また、再発リスクが高いと判断されれば、密な経過観察を行い、再発の時点で早期に治療介入することが可能となる。このように個別化した介入が可能となり、医療費削減効果が期待できる。

3) 特にPSAが一時的に上昇するPSA Bounceでは、再発か非常に判断が難しく、しばしばホルモン療法が施行されることも多いが、予後予測アルゴリズムで再発の確率が低いと判断されれば、無駄にホルモン療法が施行することがなくなる可能性がある。

本研究で明らかとした予後予測モデルが、小線源療法のみならず、他の前立腺癌放射線治療モダリティの予後予測にも応用できるかを今後明らかとしていきたい。

F . 健康危険情報

本研究はすでにJPOPSで登録され、匿名化された既存データのみを用いる観察研究であり、患者への侵襲は伴わない。

G . 研究発表

1. 論文発表

- 1) Koga H, Naito S, Ishiyama H, Yorozu A, Saito S, Kojima S, Higashide S, Kikuchi T, Nakamura K, Dokiya T, Fukushima M; J-POPS Study Group. Patient-reported health-related Quality of Life Up to Three Years after The Treatment with Permanent brachytherapy: Outcome of the large scale, prospective longitudinal study in Japanese-Prostate Cancer Outcome Study by Permanent I-125 Seed Implantation (J-POPS) brachytherapy. *Brachytherapy*. 8(6):806-813, 2019.
- 2) Huh SJ, Nishimura T, Park W, Nakamura K. Comparison of radiotherapy infrastructure between Korea and Japan. *Jpn J Clin Oncol*. 49(11):1024-1028, 2019.
- 3) Katayama N, Yorozu A, Nakamura K, Fukushima M, Kikuch T, Saito S, Dokiya T. Biochemical outcomes and predictive factors by risk group after permanent iodine-125 seed implantation: Prospective cohort study in 2,316 patients. *Brachytherapy* 18:574-582, 2019.
- 4) 中村和正. 合併症・QOL からみた各種限局がん治療の比較. 外照射療法. 泌尿器外科. 32:1085-1087, 2019

- 5) Tanaka N, Fukushima M, Yorozu A, Saito S, Ohashi T, Katayama N, Dokiya T, Kikuch T, Nakamura K, Higashide S, Kojima S. Genitourinary Toxicity after Permanent Iodine-125 Seed Implantation: The nationwide Japanese Prostate Cancer Outcome Study of Permanent Iodine-125 Seed Implantation (J-POPS). *Brachytherapy* 18:484-492, 2019.
- 6) Nakamura K, Konishi K, Komatsu T, Ishiba R. Quality of life after external beam radiotherapy for localized prostate cancer: comparison with other modalities. *Int J Urol* 26(10): 950-954, 2019.
- 7) Nakamura K, Ohga S, Yorozu A, Saito S, Kikuchi T, Dokiya T, Fukushima M, Yamanaka H. Institutional patient accrual volume and the treatment quality of I-125 prostate seed implantation in a Japanese nationwide prospective cohort study. *Strahlenther Onkol*. 195:412-419, 2019.

2. 学会発表

- 1) 中村和正、小西憲太、石場領、池之平勉、朝生智之、小松哲也. 「機械学習・汎用知的音声合成システムを利用した乳癌術後照射説明資料作成と働き方改革への寄与」日本医学放射線学会第167回中部地方会 2020.2.15-16 名古屋
- 2) 中村和正. 「ビッグデータからの機械学習による前立腺癌小線源療法の予後予

- 測法の開発と均てん化への応用 -進捗状況報告-」 第14回J-POPS中間報告会プログラム 2020.1.19 東京
- 3) 中村和正. 「前立腺癌に対する放射線治療の現状と今後の新展開」(特別講演) 第12回中部放射線医療技術学術大会 2019.11.30 浜松市
- 4) 中村和正. 「前立腺がん、放射線治療でできること」(市民公開講座)男性がん総合フォーラム2019 2019.11.24 東京
- 5) 中村和正. 「放射線腫瘍医は今後どのようにAIを利用すべきかー JPOPS研究から学んだことー」(特別企画 3 近未来の放射線治療 -AI とともに)日本放射線腫瘍学会第32回学術大会 2019.11.21-23 名古屋
- 6) Nakamura K. Recent topics in external beam radiation therapy for prostate cancer. (invited lecture) 1st Training Course of Multidisciplinary Treatment of Urinary tumors. 2019 .11.1 Beijing
- 7) 中村和正. 「前立腺癌外部照射の現状と展望」シンポジウム6 前立腺癌の低侵襲性治療 現状と展望 第55回日本医学放射線学会秋季臨床大会 2019.10.20 名古屋
- 8) Nakamura K, Magome T, Yorozu A, Saito S, Ito K, Kojima S, Kikuchi T, Fukushima M, Dokiya T, Yamanaka H. A new approach to the risk classification of patients with prostate cancer treated with I-125 prostate seed implantation in a Japanese nationwide prospective cohort study. The 61th Annual Meeting of the American Society for Radiation Oncology, September 15 - 18, 2019, Chicago, USA.
- 9) 中村和正. 「転移の早期発見・治療のために放射線のできること」前立腺がんセミナー in ジャパンキャンサーフォーラム(市民講座) 東京 2019.8.18
- 10) 中村和正. 「前立腺がんに対する放射線治療の現状と将来展望 -oligomet prostate cancerに対する放射線療法の意義を含めて-」第129回日本泌尿器科学会山陰地方会イーブニングセミナー 2019.06.22 米子
- 11) 中村和正. 「遠隔転移を有する前立腺癌の局所根治的放射線治療」シンポジウム「放射線治療の適応拡大に向けた新たな展開」第78回日本医学放射線学会総会 横浜 2019.4.12-14 (4.14)
- H . 知的財産権の出願・登録状況 (予定を含む)**
1. 特許取得
なし
 2. 実用新案登録
なし
 3. その他
なし