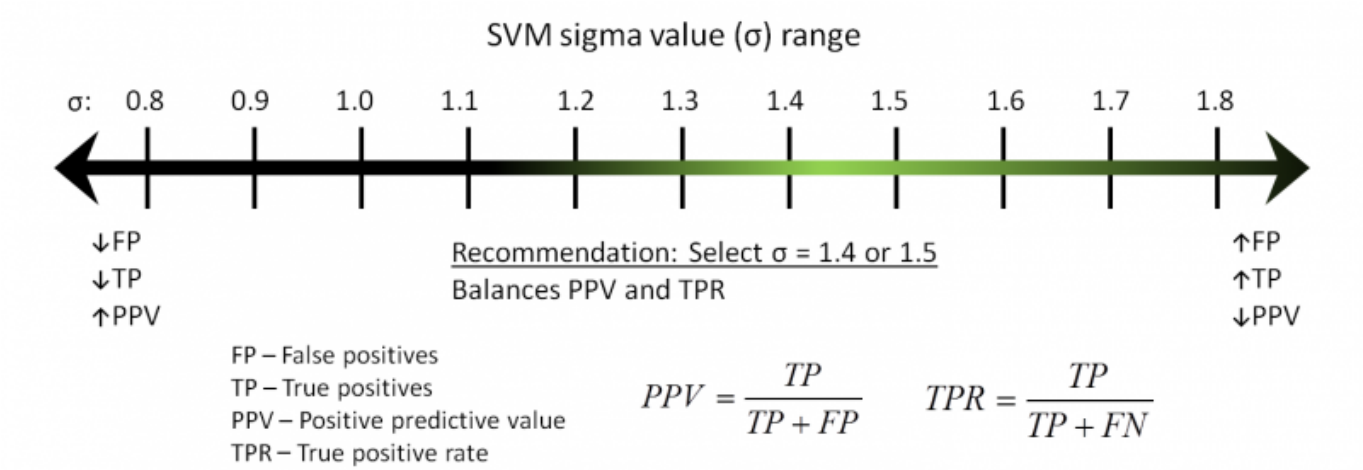
SVM sigma value

The SVM sigma value, referred to as “σ”, is an SVM tuning parameter and effectively represents a tradeoff between increased sensitivity and reduced specificity in dicentric chromosome (DC) detection. Simply put, a higher SVM sigma value will generally find more dicentric chromosomes but will result in more false positives (FPs). A lower SVM sigma value will generally result in a lower FP rate but may miss true positives (TPs). This trade-off is depicted in the image below.



The indented text and tables in this page are taken from Li et al. 2016 [1)] and modified slightly for wording and term definition only. More information about SVM sigma values can be found in the cited paper.

> There is a tradeoff between tuning the SVM to maximize either true positive rates (TPR) or positive predictive values (PPV) but not both. Increasing σ improves sensitivity, i.e., more positive predictions of DCs, but reduces specificity. However, the numbers of monocentric chromosomes (MCs) will always exceed DCs, regardless of radiation exposure. For this reason, the SVMs have been optimized to maximize correct detection of TP. σ values from 1.4 to 1.6 result in a balance of TPs and FPs and maximize PPV and TPR. At high doses, at least, these sigma values provide satisfactory accuracy for differentiating MCs from DCs, though manual review by experts is more accurate when scoring is consistent.

The tables below show the TPs, FPs, PPV, TPR in three datasets at a variant of σ values.

TABLE 2. Results of MC-DC SVM cross-validation on dataset 1.

Sigma	TPs	FPs	PPV%	TPR%[a]	TPR%[b]
1.0	91	18	83.5	46.9	34.2
1.1	111	24	82.2	57.2	41.7
1.2	124	28	81.6	63.9	46.6
1.3	134	35	79.3	69.0	50.4
1.4	148	41	78.3	76.3	55.6
1.5	154	49	75.9	79.4	57.9
2.0	166	79	67.8	85.6	62.4

[a]Total of 371 chromosomes with both centromeres correctly detected by Centromere SVM.
[b]Total of 531 chromosomes with all known DCs scored, regardless of results of Centromere SVM.

TABLE 3. Results of MC-DC SVM test on dataset 2.

σ Value	No. TPs	No. FPs	PPV%	TPR%[a]	TPR%[b]
0.9	173	65	72.6	46.6	32.6
1.0	210	96	68.6	56.6	39.6
1.1	240	149	61.7	64.7	45.2
1.2	264	186	58.7	71.2	49.7
1.3	279	234	54.4	75.2	52.5
1.4	286	264	52.0	77.1	53.9
1.5	294	302	49.3	79.3	55.4

[a]Total of 194 with both centromeres correctly detected by Centromere SVM.
[b]Total of 266 with all known DCs scored, regardless of results of Centromere SVM.

TABLE 4. Performance of MC-DC SVM on dataset 3 at different exposure levels: Comparison with expert scoring.

Source of dicentric chromosome data	Performanc criteria	SVM σ value 1.2	1.3	1.4	1.5	1.6	1.7	1.8	HC[b]	CNL[b]	UWO[a]
High-Dose chromosome data, commonly scored[b]	TPs	71	79	90	98	102	108	110	175	176	179
	FPs	13	17	33	39	46	54	66	4	3	0
	PPV%	84.5	82.3	73.2	71.5	68.9	66.7	62.5	97.8	98.3	100
	TPR%	39.7	44.1	50.3	54.8	57.0	60.3	61.5	97.8	98.3	100
All High-Dose chromosome data[c]	TPs	214	250	280	301	314	327	333	N/A	N/A	476
	FPs	43	53	81	104	125	148	172			0
	PPV%	83.3	82.5	77.6	74.3	71.5	68.8	65.9			100
	TPR%	45.0	52.5	58.8	63.2	66.0	68.7	70.0			100
Low-Dose chromosome data[c]	TPs	13	18	18	20	20	20	20	N/A	N/A	27
	FPs	37	51	67	90	120	136	156			0
	PPV%	26.0	26.1	21.2	18.2	14.3	12.8	11.4			100
	TPR%	48.2	66.7	66.7	74.1	74.1	74.1	74.1			100

[a]Results scored by University of Western Ontario (UWO/JHMK). DCs scored by UWO are treated as ground truth. Calculation of TPs and FPs based on comparing scoring by SVMs, by HC, and by CNL with ground truth.
[b]The DC chromosome subset commonly scored by UWO, Health Canada (HC), and Canadian Nuclear Laboratories (CNL) and by the software was exposed to high dose radiation.
[c]All data in the high-dose subset, scored by UWO and the software. This includes images that were not scored by all three experts. N/A, not applicable; TPs, true positives; FPs, false positive DCs; PPV, positive predictive value; TPR, true positive rate.

1)

Li Y, Knoll JHM, Wilkins R, Flegal FN, and Rogan PK. Automated Discrimination of Dicentric and Monocentric Chromosomes by Machine Learning-based Image Processing. Microscopy Research & Technique 79:393-402 (2016).