

Patching Updates

Veer Bathwal

Table of Contents

Patch Number Distribution

Epithelium Coverage

Patch Length Distribution

Outlying Cases

Patch Number Distribution

Quartiles

54

First Quartile

97

Median

155

Third Quartile

Other Statistics

1

Minimum
Case 16, match 1 (melan)

73,162

Total Patches

575

Maximum
Case 15, match 1 (h&e)

Class-Level Data

50,338

High-Grade Patches

73,162

Total Patches
(includes one case each of Class 0 and Class 2)

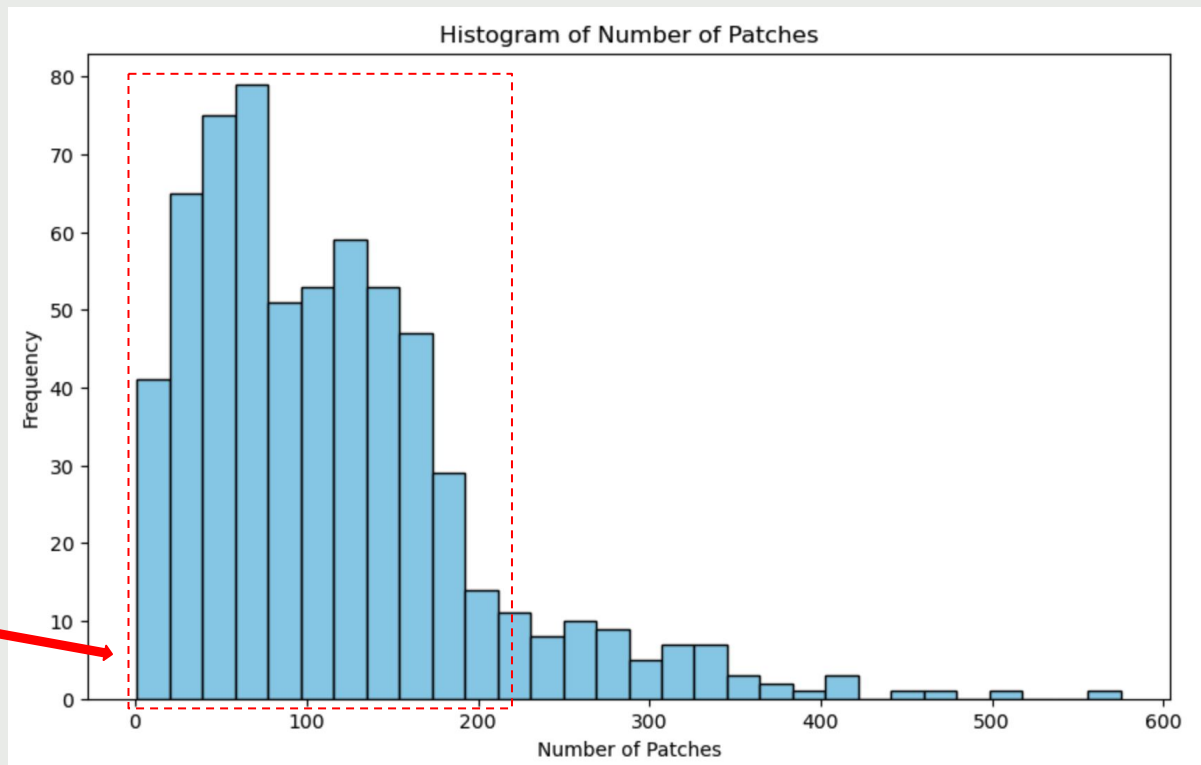
22,339

Benign Patches

Histogram

Most slices have fewer than 200 patches. We see a few outliers on the extremes, with a minimum of 1 and maximum of 575

Case 16, match 1
(melan) → 1 patch



Epithelium Coverage

Quartiles

94.57

First Quartile

96.48

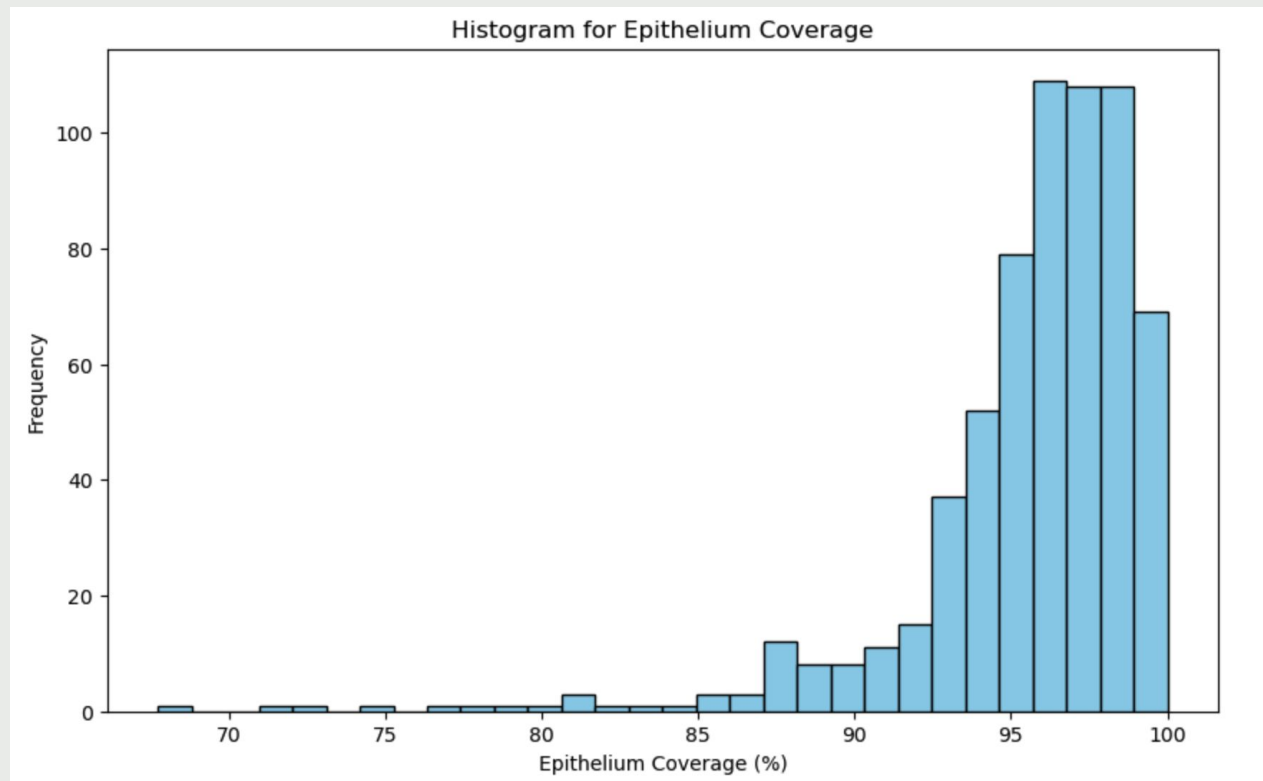
Median

97.98

Third Quartile

Histogram

Left-skewed distribution with the majority of observations greater than 95%. With the first and third quartiles being very close to the median, we have covered most of the epithelium across all slices, barring a few outliers



Patch Length Distribution

Quartiles

124

First Quartile

188

Median

276

Third Quartile

Other Statistics

4

Minimum

2048

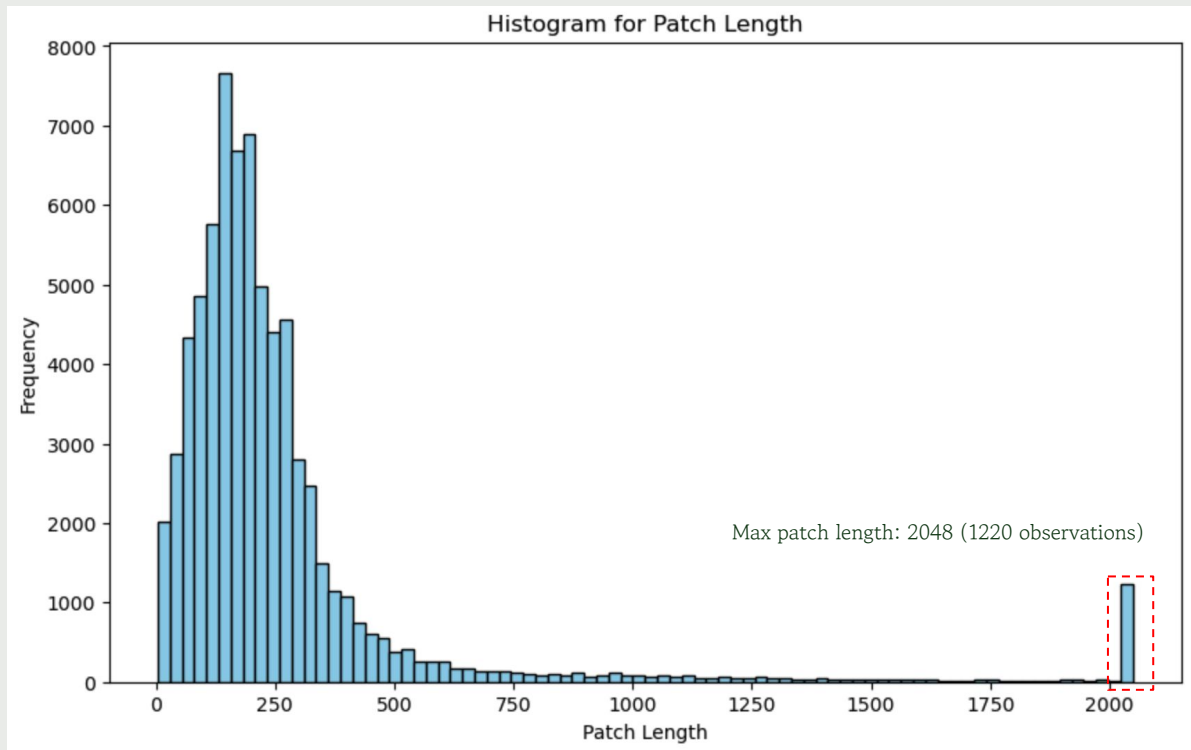
Mode (1220 Occurrences)

2048

Maximum

Histogram

Right-skewed distribution, except for a peak at max length of 2048, which is also the mode (as described in patching algorithm). Majority of the observations fall below the 250 mark (close to the third quartile value of 280)



Outlying Cases

General Observations

136

Number of slices with <50 patches

14

Number of slices with <85% epithelium coverage

395

Number of patches with length < 10

Num Patches < 50

We see 136 slices with fewer than 50 patches

```
metrics_df[metrics_df['num_patches']<50]
```

	slice	num_patches	epithelium_coverage	case	Class
9	case_28_unmatched_1_h&e	26.0	97.851950	28	1
11	case_40_match_3_h&e	42.0	99.790971	40	4
16	case_63_match_2_h&e	49.0	96.607591	63	3
26	case_38_match14_sox10	17.0	99.914552	38	4
30	case_070_match_6_melan	37.0	99.632820	70	4
...
629	case_14_match_3_sox10	45.0	99.302244	14	4
630	case_42_match_2_h&e	28.0	93.258709	42	3
631	case_63_match_7_h&e	10.0	91.417216	63	3
633	case_88_unmatched2_h&e	19.0	94.213062	88	4
635	case_49_match_1_sox10	46.0	98.958394	49	4

136 rows × 5 columns

```
metrics_df['num_patches'].nsmallest(10)
```

70 1.0

128 1.0

617 4.0

374 5.0

508 5.0

527 5.0

204 9.0

268 10.0

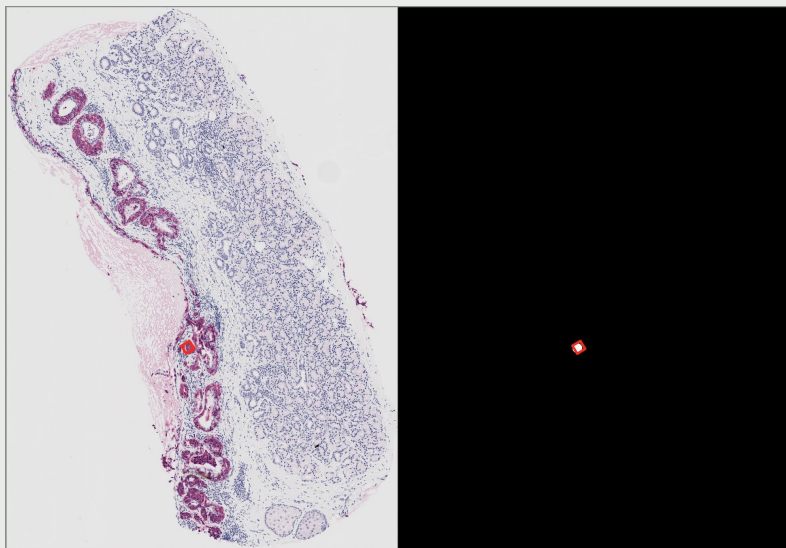
376 10.0

460 10.0

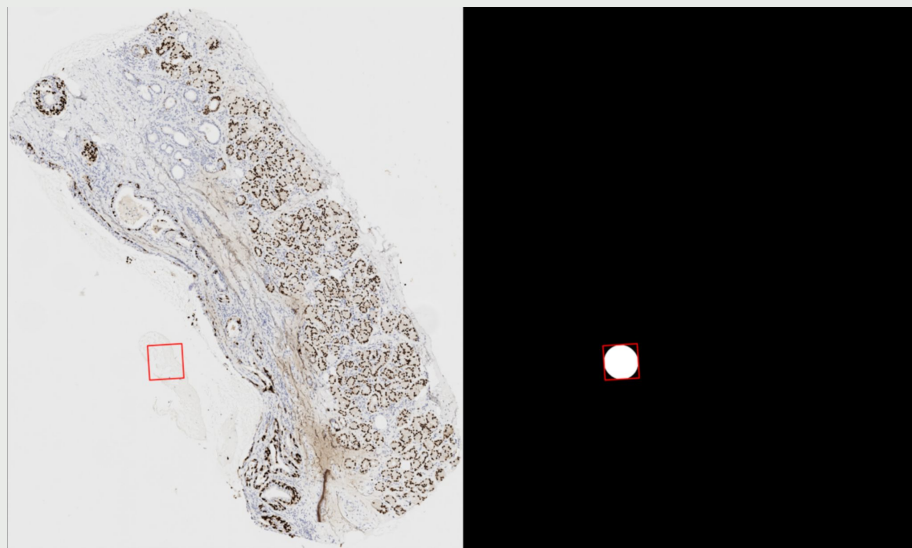
Name: num_patches, dtype: float64

Case 16, Match 1

Slice annotation issue for melan and sox10 slices – problem not observed in h&e stain

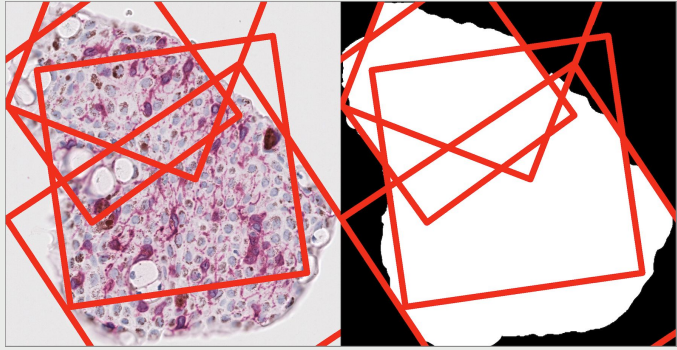


Melan

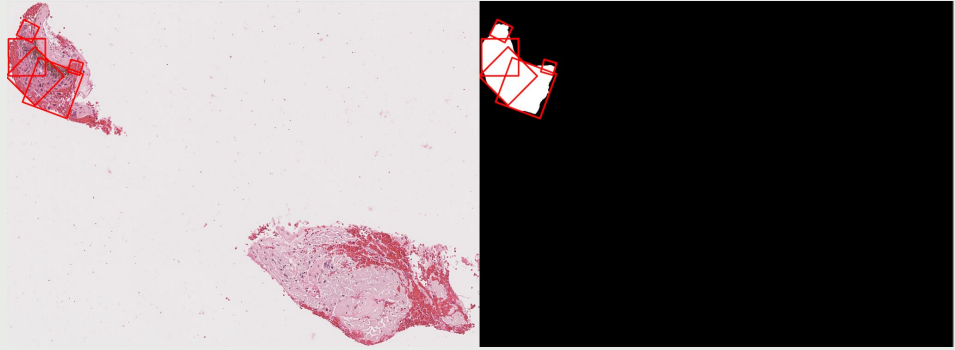


Sox10

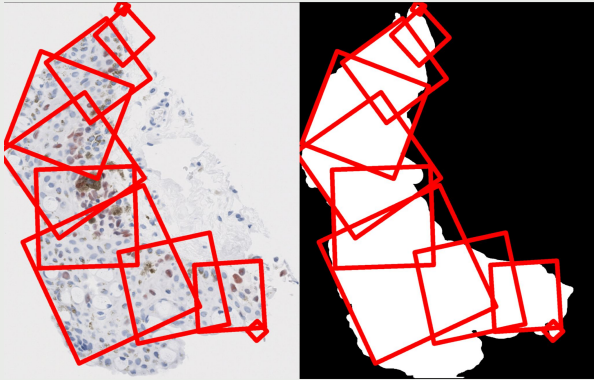
Other slices with extremely low num patches



Case 63, Match 6 (Melan and Sox10)



Case 72, Unmatched 4 (h&e)



Case 69, Match 3 (h&e)

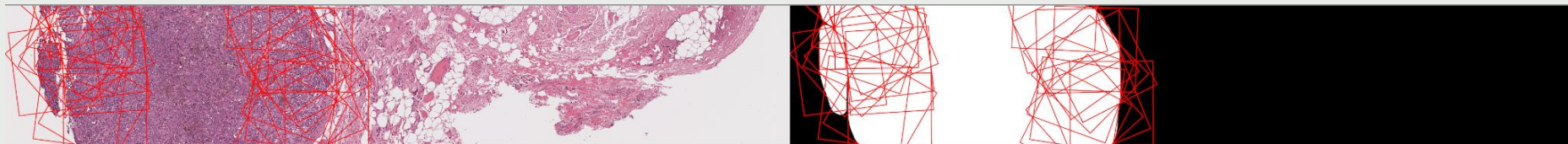
Epithelium Coverage < 85%

We see 14 slices falling into this category. Most of them seem concentrated in a few particular cases. Notably, case 38 and 50.

```
metrics_df[metrics_df['epithelium_coverage']<85]
```

	slice	num_patches	epithelium_coverage	case	Class
76	case_066_match_4_melan	150.0	84.589035	66	3
98	case_50_match_4_melan	40.0	79.066646	50	3
175	case_50_match_3_melan	334.0	82.624301	50	3
202	case_19_match_2_melan	152.0	72.037059	19	3
236	case_38_match_27_melan	27.0	73.091255	38	4
241	case_38_match13_h&e	31.0	67.737730	38	4
257	case_38_match_31_h&e	46.0	80.717639	38	4
267	case_50_match_3_sox10	237.0	81.711564	50	3
280	case_066_match_3_melan	94.0	80.597726	66	3
392	case_038_match_11_melan	54.0	80.702881	38	4
439	case_38_match_29_sox10	28.0	77.735912	38	4
456	case_038_match_11_sox10	30.0	74.200038	38	4
531	case_4_unmatched_1_h&e	163.0	83.356913	4	1
622	case_85_unmatched2_h&e	76.0	76.936591	85	1

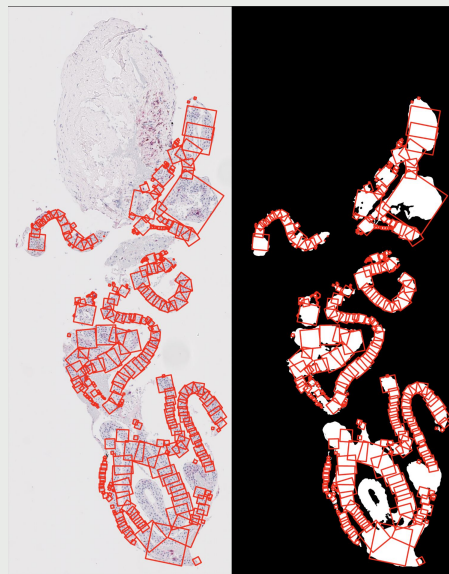
Samples



Case 38, Match 13 (h&e)



Case 19, Match 2 (melan)



Case 50, Match 3 (melan)

Takeaways

- Other than the outliers, patches seem to be well-formed and substantially cover the epithelium for the most part
- However, slices from cases such as 16, 63, 38, and 50 to name a few that commonly pop up in the list of outliers should be left out from our models for now, to prevent feeding in bad quality training data
- A lot of the issues still persist due to bad annotations
- Should consider trade-off of spending another week fixing those patches (by re-annotating) versus completely moving focus to model-building. Is the number of “bad” slices high enough to allocate more time toward patching?

Thank You