

Contents

Overall SEER JSON Structure	2
Prediction Output Structure Explained	3
Color -WSI-MultiRes explained	3
Prediction -WSI-MultiRes explained	4
Prediction -WSI-MultiRes for low resolution images explained.....	5
Cluster Output Structure Explained.....	6
Overall cluster statistics output.....	6
Overall cluster summary information	7
Cluster Indices output (AP algorithm)	8

Overall SEER JSON Structure

The complete JSON file is available here in Github. [SEER.json](#)

The diagram illustrates the structure of the SEER JSON file, which is a single object containing several key-value pairs. The JSON is displayed on the left, with line numbers 1 through 270. Three yellow boxes highlight specific sections of the JSON, with arrows pointing to explanatory text boxes on the right.

```
1 {  
2   "_id": "5bab78756ecaeab04c649c29",  
3   "cancer_type": "SEER",  
4   "WSI-name": "N3908-HE-CROP",  
5   "prediction-step-output": {  
6     "color-WSI-multires": [ ... ],  
21    ],  
22    "prediction-WSI-multires": [ ... ],  
35    ],  
36    "prediction-WSI-multires.low_res": [ ... ],  
49    ],  
50  },  
51  "cluster-step-output": {  
52    "seer_totalClusterStats": [ ... ],  
147   ],  
148    "WSI-multires_clusterInfo": { ... },  
152   },  
153    "WSI-multires_indices_ap": [ ... ],  
268   ],  
269  },  
270 }
```

Summary of a WSI :

1. cancer type
2. name of WSI

CSVs generated from predictions by the CNN Model

1. multiple resolution color(whiteness, redness, blackness) measures
2. Prediction on high resolution images for every coordinate of patches
3. Prediction on low resolution images for every coordinate of patches

CSVs generated from clustering step (AP algorithm output)

1. Overall clustering statistics on the WSI
2. Cluster Summary Information for the WSI
3. Cluster Indices output from the AP algorithm run on the WSI

Prediction Output Structure Explained


Color -WSI-MultiRes explained

```
"prediction-step-output": {  
  "color-WSI-multires": [  
    {  
      "coor_x": 1,  
      "coor_y": 1,  
      "whiteness": 0,  
      "blackness": 0,  
      "redness": 0  
    },  
    {  
      "coor_x": 1,  
      "coor_y": 2,  
      "whiteness": 0,  
      "blackness": 0,  
      "redness": 0  
    }  
  ],  
}
```

For every x and y coordinate of the patches, what is the color density information based on the predictions produced by the CNN model. This will be useful to generate the PNG images

Note: Only 2 data points have been shown for illustration purposes. Obviously in actual data all coordinates of the image will be available.

Prediction -WSI-MultiRes explained

```
"prediction-step-output": {  
  "color-WSI-multires": [ ,  
  ],  
  "prediction-WSI-multires": [  
    {  
      "patch_index_x": 99,  
      "patch_index_y": 99,  
      "bin_value": 0,  
      "real_value": 0.00722075533122  
    },  
    {  
      "patch_index_x": 99,  
      "patch_index_y": 296,  
      "bin_value": 0,  
      "real_value": 0.00555958598852  
    }  
  ],  
}
```

For each coordinate of x and y in the patch, this section captures the following

bin_value: TIL/No-TIL (1/0)

real_value: Actual Probability predicted value

Prediction -WSI-MultiRes for low resolution images explained

```
"prediction-step-output": {  
  "color-WSI-multires": [ =  
  ],  
  "prediction-WSI-multires": [ =  
  ],  
  "prediction-WSI-multires.low_res": [  
    {  
      "patch_index_x": 392,  
      "patch_index_y": 392,  
      "bin_value": 0,  
      "real_value": 0.004871  
    },  
    {  
      "patch_index_x": 392,  
      "patch_index_y": 1176,  
      "bin_value": 0,  
      "real_value": 0.005726  
    }  
  ]  
}
```

For low resolution images in each coordinate of x and y, this section conveys the following information:

1. bin_value: Til/No Til (1/0)
2. real_value: Actual Probability value

Cluster Output Structure Explained

Overall cluster statistics output

```
"cluster-step-output": {  
  "seer_totalClusterStats": [  
    {  
      "Slides": "N3908-HE-CROP-multires.csv",  
      "number of data points": 363,  
      "number of clusters": 4,  
      "Ball_Hall": 727.033492228104,  
      "Banfeld_Raftery": 2335.98088004097,  
      "C_index": 0.0335152159541665,  
      "Calinski_Harabasz": 968.214411749257,  
      "Davies_Bouldin": 0.430478646107358,  
      "Det_Ratio": 59.3988774907265,  
      "Dunn": 0.052702797742987,  
      "Gamma": 0.917867183685303,  
      "G_plus": 0.0164737177128835,  
      "GDI11": 0.052702797742987,  
      "GDI12": 0.339330435992321,  
      "GDI13": 0.108306441285547,  
      "GDI21": 1.28806866834899,  
    }  
  ]  
}
```

This is the first section of the clustering output.

This section represents the overall cluster statistics for each WSI. Some of the fields have been listed, the actual JSON has all the fields available as part of this statistics.

Overall cluster summary information

```
"cluster-step-output": {  
  "seer_totalClusterStats": [  
    ],  
    "WSI-multires_clusterInfo": {  
      "Slides": "N3908-HE-CROP-multires",  
      "number of data points": 363,  
      "number of clusters": 4  
    },  
  ],  
}
```

This section represents the general cluster information for the WSI.

1. Number of data points
2. Number of clusters detected by AP algorithm

Cluster Indices output (AP algorithm)

```
51 ▾ "cluster-step-output": {  
52 ▸   "seer_totalClusterStats": [▮  
147   ],  
148 ▸   "WSI-multires_clusterInfo": {▮  
152   },  
153 ▾ "WSI-multires_indices_ap": [  
154 ▾   {  
155     "Slides": "TCGA-05-4396-01Z-00-DX1",  
156     "Ball_Hall": "NA",  
157     "Banfeld_Raftery": "NA",  
158     "C_index": "NA",  
159     "Calinski_Harabasz": "NA",  
160     "Davies_Bouldin": "NA",  
161     "Det_Ratio": "NA",  
162     "Dunn": "NA",  
163     "Gamma": "NA",  
164     "G_plus": "NA",  
165     "GDI11": "NA",  
166     "GDI12": "NA",  
167     "GDI13": "NA",  
168     "GDI21": "NA",  
169     "GDI22": "NA",  
170     "GDI23": "NA",  
171     "GDI31": "NA",  
172     "GDI32": "NA",  
173     "GDI33": "NA",  
174     "GDI41": "NA",  
175     "GDI42": "NA",
```

This section captures the output as part of the clustering indexing step by the AP algorithm:

It is an array of data elements (a subset has been shown for illustration purposes).