

## Contents

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

## Overall SEER JSON Structure

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

The diagram illustrates the structure of a SEER JSON file, showing a code editor on the left and explanatory text boxes on the right. The JSON is a single object with three main properties: `_id`, `cancer_type`, and `WSI-name`. The `prediction-step-output` property contains three arrays: `color-WSI-multires`, `prediction-WSI-multires`, and `prediction-WSI-multires.low_res`. The `cluster-step-output` property contains three objects: `seer_totalClusterStats`, `WSI-multires_clusterInfo`, and `WSI-multires_indices_ap`.

**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

```
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 }
```

## Prediction Output Structure Explained

These outputs reside in the following folder structure:

<TIL root>/data/heatmap\_txt/

<TIL root>/cluster\_indices/input/seer

### Color -WSI-MultiRes explained

It tells whether a portion of the image is background or tissue. It is statistically computed, not generated by the model.

```
"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 and low resolution explained

```
6  "prediction-step-output": {
7    "color-WSI-multires": [
8      { =
14    },
15    { =
21    }
22  ],
23  "prediction-WSI-multires": [
24    {
25      "patch_index_x": 99,
26      "patch_index_y": 99,
27      "lymphocyte_pred": 0,
28      "necrosis_pred": 0.00722075533122
29    },
30    {
31      "patch_index_x": 99,
32      "patch_index_y": 296,
33      "lymphocyte_pred": 0,
34      "necrosis_pred": 0.00555958598852
35    }
36  ],
37  "prediction-WSI-multires.low_res": [
38    {
39      "patch_index_x": 392,
40      "patch_index_y": 392,
41      "lymphocyte_pred": 0,
42      "necrosis_pred": 0.004871
43    }
44  ]
45 }
```

Each unique combination of X and Y value in every record represents a 100 X 100 pixels patch.

The corresponding TIL/Lymphocyte prediction and Necrosis Prediction is given for every patch once the CNN prediction model has executed and generated these outputs.

## Cluster Output Structure Explained

Cluster Index Step CSV outputs are available here:

<TIL root>/cluster\_indices/output/<cancer-type>

<TIL root>/cluster\_indices/output/seer

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).