Outcome of MOGAE on Melanoma dataset

The Multi-Objective Genetic Algorithm Explainer (MOGAE) is a model-agnostic local explainer. By model-agnostic, we are referring to the explainer's ability to explain any learning models. In the local context, the focus is on explaining individual samples. In our corresponding paper, we demonstrated MOGAE's capability in explaining the fine-tuned ResNet50 for predicting citrus diseases in plants using their leaves. In this study, we also explored the results of MOGAE on another ResNet50 model trained to predict melanoma images. We specifically selected 8 images, as illustrated in Figure 1, belonging to three categories of pigmented skin lesions. The clinician delineations of these images are provided in Figure 2. These images were also subjected to explanation using the Ensemble-based Genetic Algorithm Explainer[1] (EGAE). EGAE incorporates 5 Genetic Algorithms (GAs), each initialized with a different number of superpixels. Both EGAE and MOGAE utilize the SLIC segmentation algorithm. The number of superpixels in the last GA of EGAE was set at 100, indicating the granularity of EGAE. For a fair comparison, we standardized the number of superpixels to 100 in MOGAE as well, while keeping the remaining setups of MOGAE unchanged, as stated in the paper.

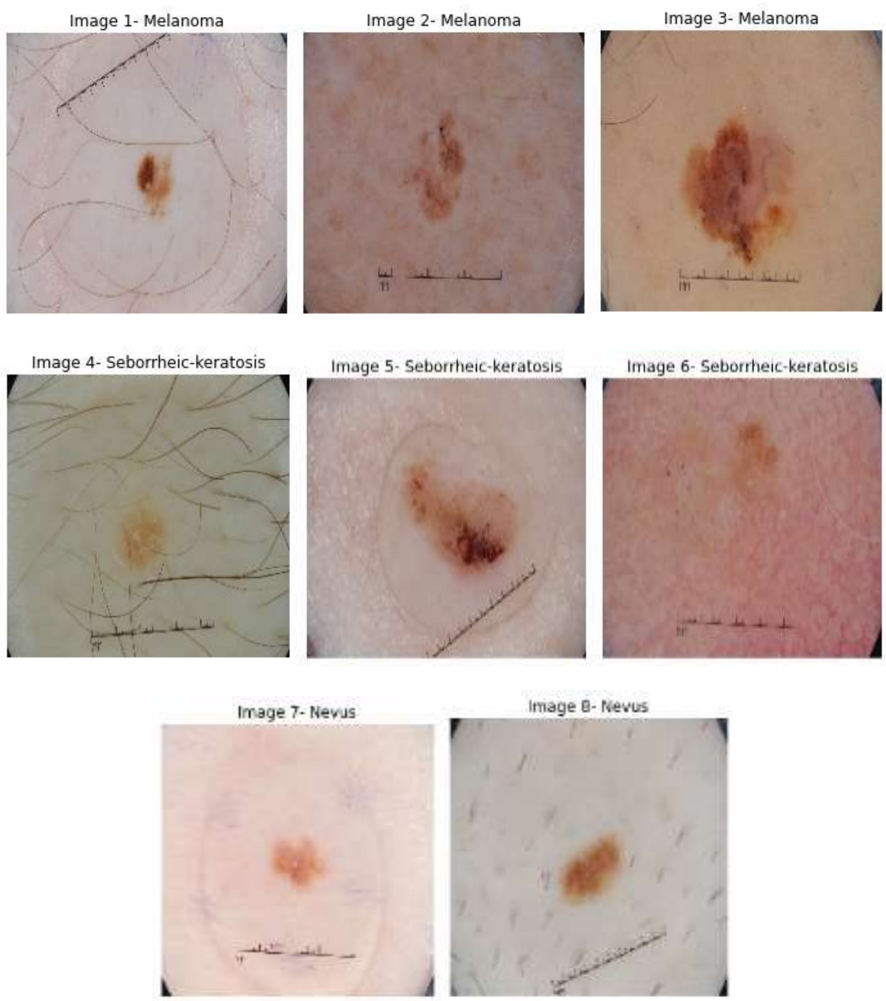

Figure 1. Illustrations of selected test data for evaluation of MOGAE.

[1] https://github.com/KhaosResearch/EGAE

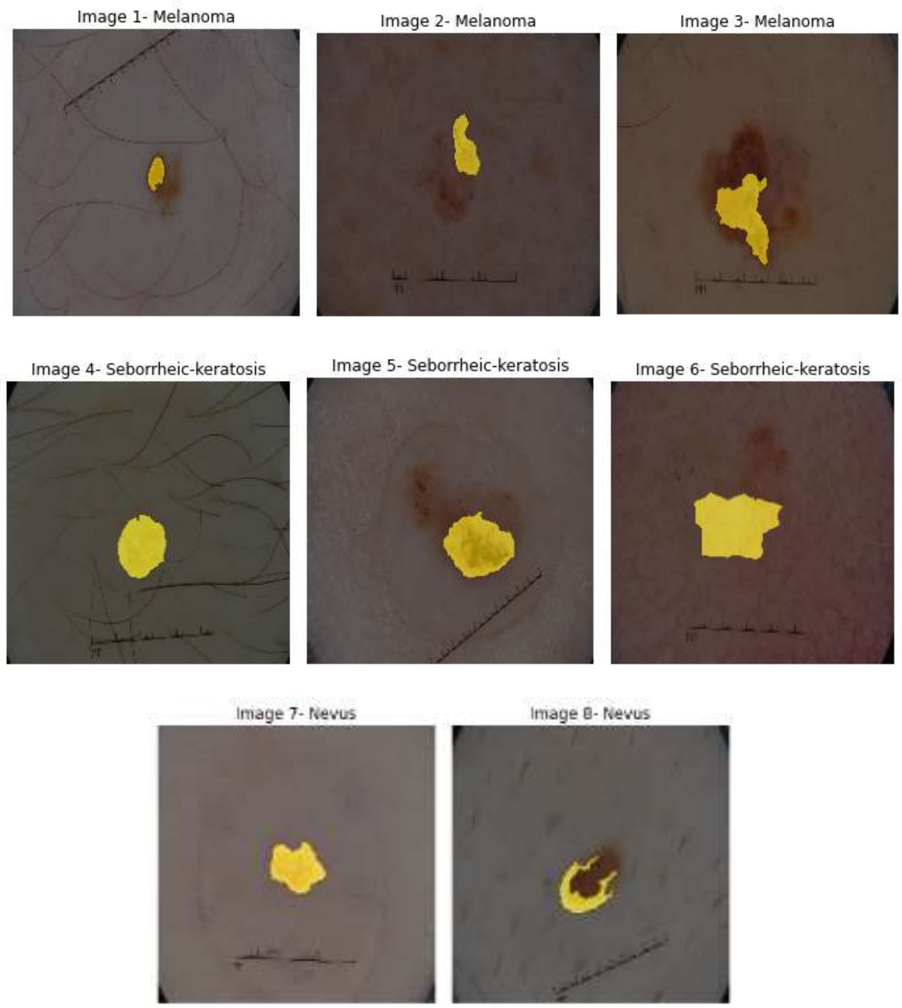

Figure 2. Delineations of the clinician generated by SLIC segmentation algorithm.

Figure 3 displays the results of MOGAE on the images from Figure 1. As MOGAE is based on NSGA-II, we executed MOGAE three times to capture the stochastic characteristics and ensure reproducibility. In addition to majority voting, we also employed consensus voting on the optimal images in the Pareto front for further investigation. Consensus voting is stricter than majority voting. It selects the regions of the image only if the optimal images in the Pareto front unanimously believe those regions are important for explanation. The strictness sometimes leads to unexpected situations, such as in the first attempt of image 3, where the result of consensus voting incorporates very few pixels. Consequently, the explanation image is hardly visible and comprehensible to the user. This issue was also observed with EGAE, and one reason might be the use of evolutionary algorithms in both methods. In some other scenarios, like the second attempt of image 5, both consensus voting and majority voting tend to produce identical explanations. However, in most scenarios, MOGAE with majority voting tends to select more regions of the image.

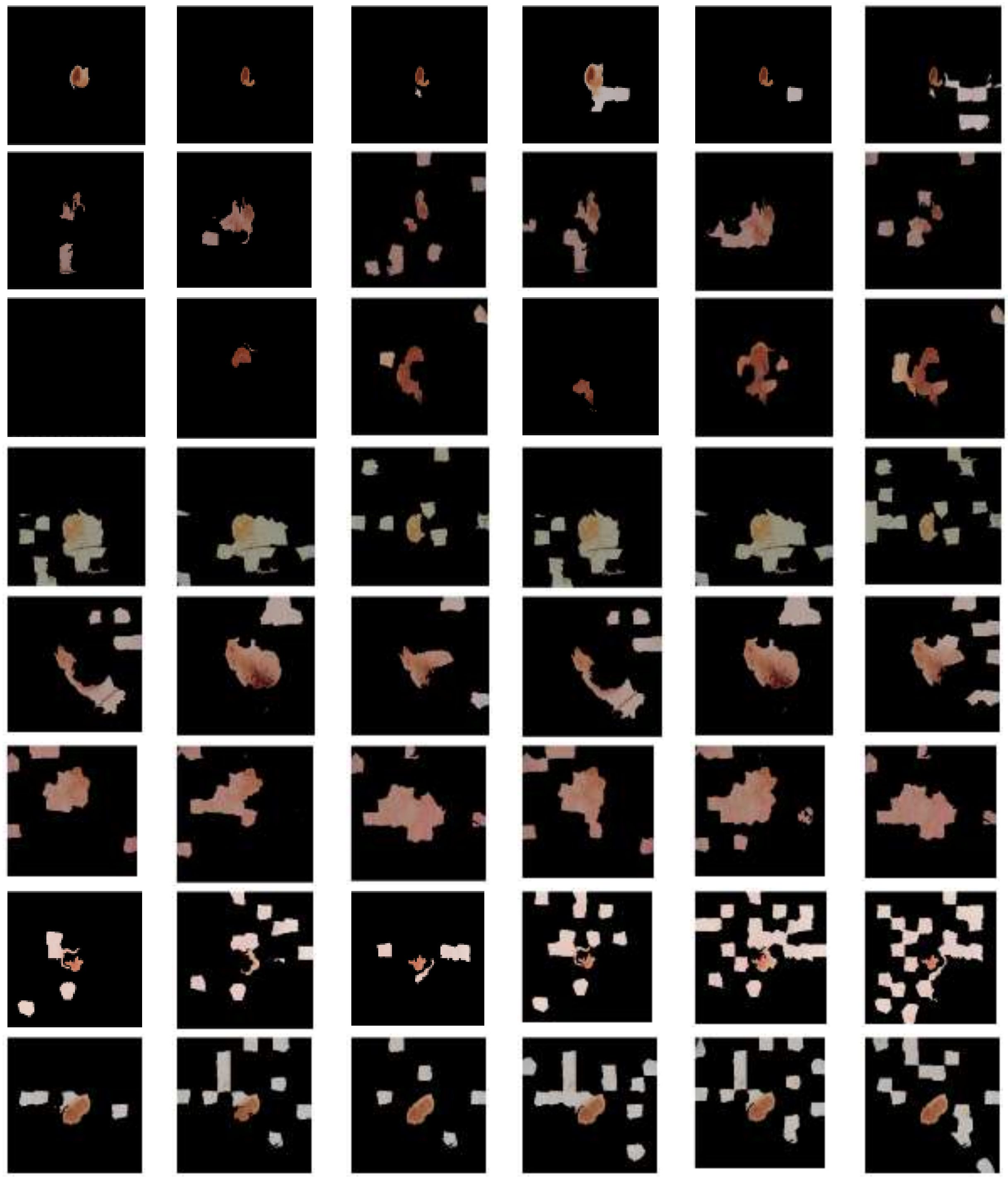

Figure 3. The results of MOGAE in three runs.

Table 1 compares the explanation errors achieved from voting approaches (consensus vs. majority) of MOGAE. As expected, the explanation image from MOGAE with consensus voting has less error than majority voting. We also compared the normalized explanation errors between MOGAE and EGAE with their consensus and majority voting strategies in Figure 4. Figure 4 demonstrates that the consensus voting strategies in either EGAE or MOGAE have less error compared to the corresponding majority voting strategies. However, EGAE with consensus voting slightly outperformed MOGAE with consensus voting in general. Nonetheless, MOGAE with majority voting slightly outperformed EGAE with majority voting generally.

Table 1. Comparison of the non-normalized explanation error of voting strategies.

methods	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8
MOGAE – consensus voting	17	41	32	68	73	75	84	80
MOGAE – majority voting	50	52	35	76	81	79	137	111

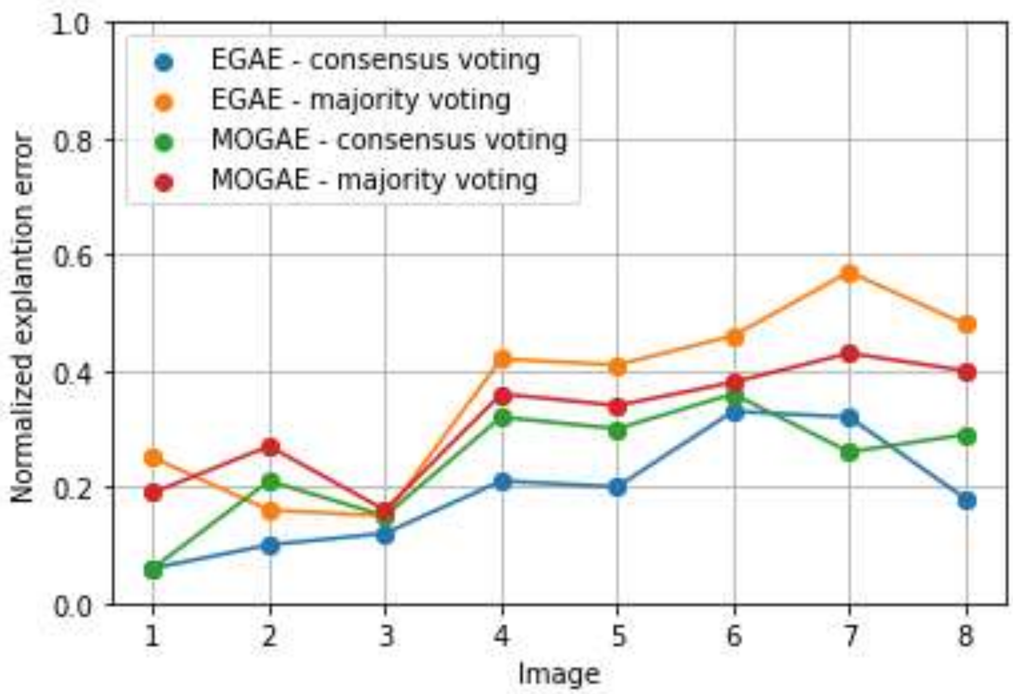

Figure 4. Comparison of MOGAE Vs. EGAE based on normalized explanation error achieved using different voting strategies.

Figure 5 also compares the Number of Function Evaluations (NFE) achieved by both EGAE and MOGAE. MOGAE adheres to a fixed number of NFE for any image; however, EGAE obtains different NFEs for different images. Considering that the voting strategies do not have any impact on different NFEs, Figure 5 clearly shows that when the number of superpixels in the segmentation algorithm is high (100 in our scenario, for example), MOGAE tends to use fewer function evaluations to generate the explanation image compared to EGAE, demonstrating a clear superiority.

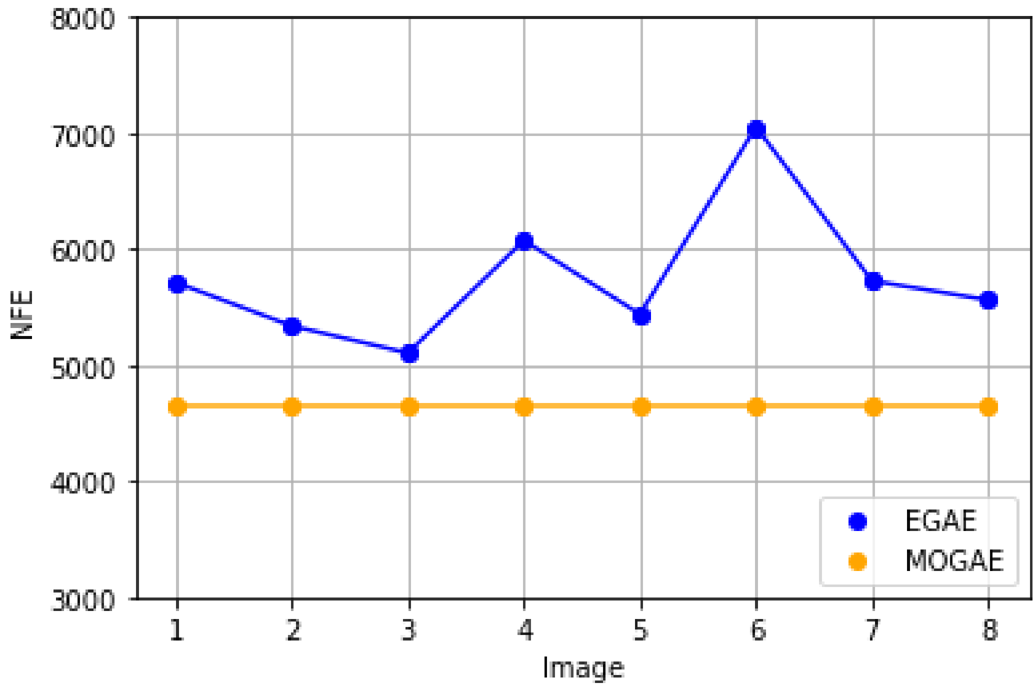

Figure 5. Comparison of MOGAE Vs. EGAE based on the Number of Function Evaluations (NFE) achieved.

All in all, EGAE and MOGAE are not distinctly separated based solely on the accuracy of explanation, as this depends on specific images and scenarios. Therefore, we suggest using both methods on different datasets to experimentally investigate the results. However, when the number of superpixels is high, MOGAE generally outperforms EGAE. This bold characteristic becomes particularly relevant when domain experts seek more details in the input image.