## A. Additional Model Training Details

The specifications for the model training have been provided in Tables 2 to Table 4.

| Hyperparams. | CIFAR-10 | FER+ | CHD |
|---|---|---|---|
| optimizer | adam | adam | RMSProp |
| loss | cross-entropy | cross-entropy | MSE |
| metric | accuracy | accuracy | MAE |
| learning rate | 0.001 | 0.001 | 0.001 |
| train samples | 50000 | 28559 | 12750 |
| validation samples | 9500 | 3579 | 4250 |
| test samples | 500 | 3573 | 4250 |
| train batch size | 2048 | 2048 | 4096 |
| validation batch size | 2048 | 2048 | 4096 |

Table 2. Hyperparameter used to train the miniVGG16 on CIFAR-10, FER+ and California Housing Dataset (CHD). MSE: Mean Squared Error, MAE: Mean Absolute Error.

| Augmentation | CIFAR-10 | FER+ |
|---|---|---|
| width shift range | 0.1 | 0.1 |
| height shift range | 0.1 | 0.1 |
| shear range | 0.1 | - |
| zoom range | 0.1 | 0.1 |
| horizontal flip | True | True |
| rescale | 1/255 | 1/255 |
| rotation range | - | 30 |

Table 3. Data augmentation applied to CIFAR-10 and FER+.

| Hyperparameter | Search Space |
|---|---|
| # of units in a layer | {1, 2, 4, 6, 8} |
| # of layers | {1, 2, 3, 4, 5} |
| # of epochs | {10, 50, 100, 150} |
| optimizer | {'SGD', 'RMSProp', 'Adam'} |

Table 4. Search space used for the hyperparameter tuning conducted for the MLP.

## B. Additional Insertion/Deletion Curves

The Area Under Curve (AUCs) for the pixel insertion and deletion metrics for CIFAR-10 and FER+ are provided in Table 5. Additionally, the class specific pixel insertion and deletion curves for FER+ are shown in Figure 10.

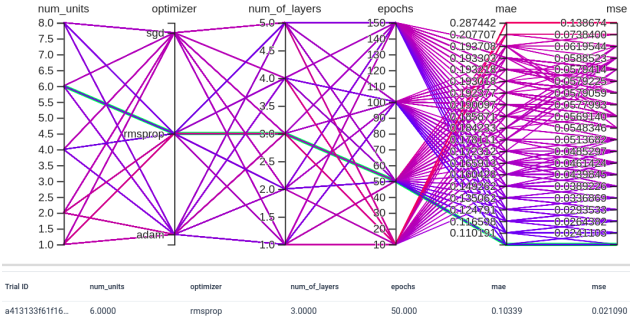


Figure 7. Parallel plot depicting the results of hyperparameter search conducted for creating an MLP. The optimum values of the hyperparameters are provided in the table below the plot.

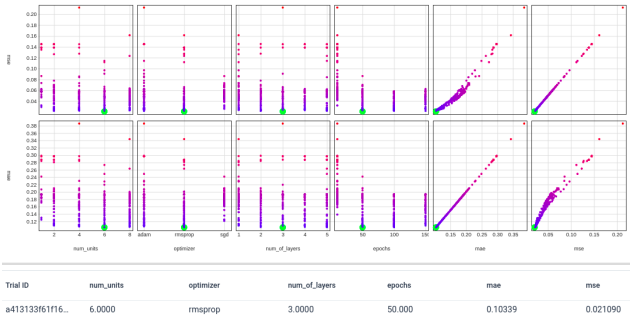


Figure 8. Scatter plot depicting the results of hyperparameter search conducted for creating an MLP. The optimum values of the hyperparameters are provided in the table below the plot.

## C. Additional Example Explanations

Additional examples of explanation uncertainty (similar to Figures 4 and 5) have been provided in Figures 11 - 16 for both the CIFAR10 and FER+ datasets. It should be noted that the heatmaps in the additional examples are *not* normalized, unlike the figures in the main paper which have proper normalization.

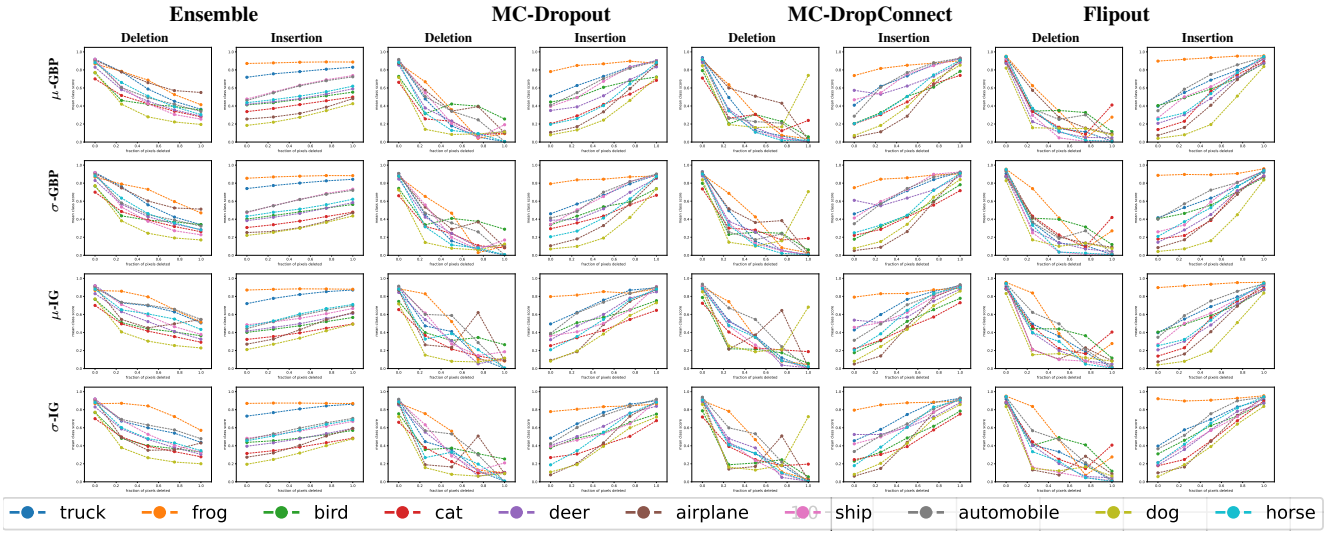


Figure 9. Pixel flipping curves for Deep Ensemble (columns 1-2), MC-Dropout (columns 3-4), MC-DropConnect (columns 5-6), and Flipout (columns 7-8) for CIFAR-10. The columns 1, 3, 5, and 7 depict the deletion curves whereas the columns 2, 4, 6, and 8 depict the insertion curves. From top to bottom, the rows show the following heatmaps and the explanation methods they correspond to: (i) $\mu$-GBP (ii) $\sigma$-GBP (iii) $\mu$-IG (iv) $\sigma$-IG.

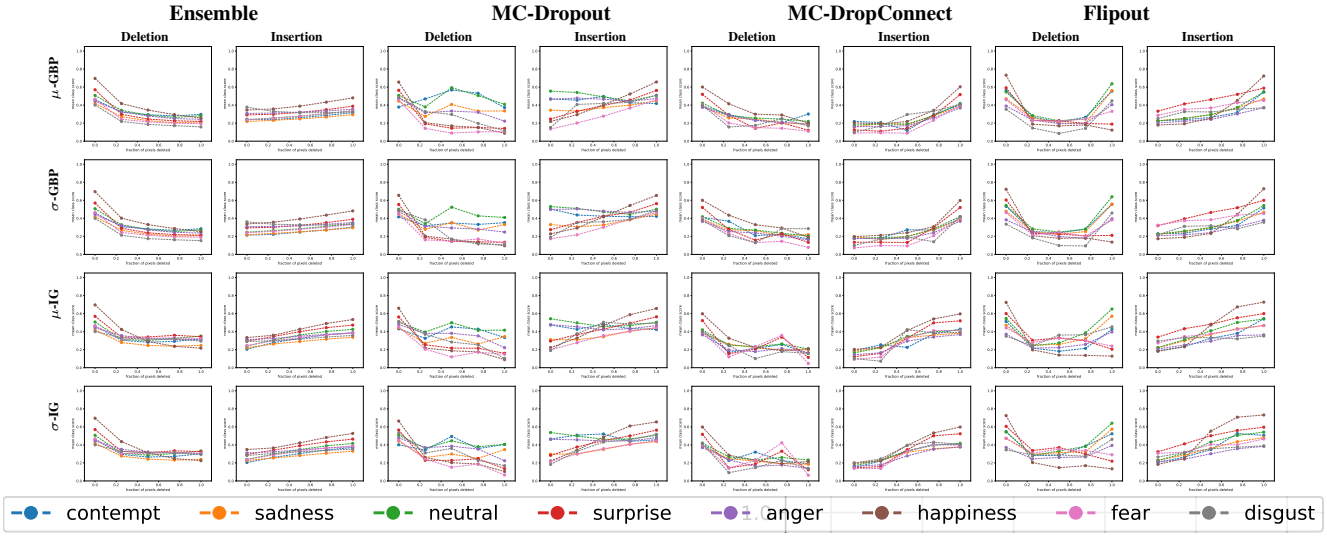


Figure 10. Pixel flipping curves for Deep Ensemble (columns 1-2), MC-Dropout (columns 3-4), MC-DropConnect (columns 5-6), and Flipout (columns 7-8) for FER+. The columns 1, 3, 5, and 7 depict the deletion curves whereas the columns 2, 4, 6, and 8 depict the insertion curves. From top to bottom, the rows show the following heatmaps and the explanation methods they correspond to: (i) $\mu$-GBP (ii) $\sigma$-GBP (iii) $\mu$-IG (iv) $\sigma$-IG.

| | CIFAR10 | | | | | | | | | | | | | | | | | | | | FER+ | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **truck** (+ -) | | **frog** (+ -) | | **bird** (+ -) | | **cat** (+ -) | | **deer** (+ -) | | **plane** (+ -) | | **ship** (+ -) | | **auto** (+ -) | | **dog** (+ -) | | **horse** (+ -) | | **contempt** (+ -) | | **sadness** (+ -) | | **neutral** (+ -) | | **surprise** (+ -) | | **anger** (+ -) | | **happy** (+ -) | | **fear** (+ -) | | **disgust** (+ -) | |
| **$\mu$-GBP-Ensemble** | 0.80 | 0.62 | 0.86 | 0.65 | 0.48 | 0.52 | 0.40 | 0.47 | 0.47 | 0.52 | 0.34 | 0.69 | 0.61 | 0.53 | 0.59 | 0.57 | 0.30 | 0.41 | 0.52 | 0.56 | 0.24 | 0.35 | 0.25 | 0.29 | 0.28 | 0.36 | 0.33 | 0.35 | 0.31 | 0.32 | 0.40 | 0.44 | 0.27 | 0.30 | 0.29 | 0.25 |
| **$\sigma$-GBP-Ensemble** | 0.78 | 0.59 | 0.87 | 0.69 | 0.47 | 0.51 | 0.40 | 0.45 | 0.45 | 0.50 | 0.34 | 0.66 | 0.62 | 0.51 | 0.59 | 0.56 | 0.31 | 0.39 | 0.51 | 0.54 | 0.25 | 0.34 | 0.25 | 0.28 | 0.28 | 0.35 | 0.33 | 0.34 | 0.32 | 0.32 | 0.40 | 0.43 | 0.27 | 0.29 | 0.31 | 0.24 |
| **$\mu$-IG-Ensemble** | 0.79 | 0.72 | 0.86 | 0.73 | 0.48 | 0.53 | 0.40 | 0.47 | 0.46 | 0.56 | 0.43 | 0.62 | 0.58 | 0.62 | 0.57 | 0.73 | 0.32 | 0.43 | 0.58 | 0.65 | 0.31 | 0.35 | 0.27 | 0.30 | 0.34 | 0.39 | 0.38 | 0.42 | 0.34 | 0.35 | 0.43 | 0.40 | 0.32 | 0.37 | 0.35 | 0.33 |
| **$\sigma$-IG-Ensemble** | 0.78 | 0.66 | 0.86 | 0.77 | 0.48 | 0.51 | 0.39 | 0.45 | 0.48 | 0.54 | 0.40 | 0.54 | 0.56 | 0.59 | 0.57 | 0.68 | 0.32 | 0.41 | 0.58 | 0.56 | 0.33 | 0.35 | 0.27 | 0.30 | 0.33 | 0.38 | 0.38 | 0.41 | 0.34 | 0.35 | 0.43 | 0.41 | 0.31 | 0.36 | 0.35 | 0.33 |
| **$\mu$-GBP-Dropout** | 0.69 | 0.29 | 0.86 | 0.39 | 0.56 | 0.39 | 0.45 | 0.25 | 0.53 | 0.29 | 0.37 | 0.45 | 0.65 | 0.33 | 0.68 | 0.38 | 0.30 | 0.19 | 0.42 | 0.24 | 0.46 | 0.38 | 0.34 | 0.37 | 0.48 | 0.48 | 0.40 | 0.21 | 0.46 | 0.33 | 0.38 | 0.22 | 0.28 | 0.11 | 0.39 | 0.24 |
| **$\sigma$-GBP-Dropout** | 0.70 | 0.29 | 0.85 | 0.40 | 0.53 | 0.40 | 0.44 | 0.27 | 0.52 | 0.32 | 0.38 | 0.41 | 0.62 | 0.34 | 0.66 | 0.38 | 0.28 | 0.19 | 0.43 | 0.23 | 0.45 | 0.36 | 0.37 | 0.33 | 0.48 | 0.43 | 0.39 | 0.21 | 0.47 | 0.31 | 0.41 | 0.22 | 0.31 | 0.18 | 0.33 | 0.21 |
| **$\mu$-IG-Dropout** | 0.75 | 0.38 | 0.82 | 0.48 | 0.54 | 0.39 | 0.43 | 0.28 | 0.61 | 0.35 | 0.47 | 0.38 | 0.56 | 0.40 | 0.69 | 0.49 | 0.39 | 0.19 | 0.56 | 0.34 | 0.46 | 0.41 | 0.36 | 0.32 | 0.49 | 0.43 | 0.40 | 0.26 | 0.46 | 0.36 | 0.45 | 0.24 | 0.30 | 0.16 | 0.45 | 0.31 |
| **$\sigma$-IG-Dropout** | 0.73 | 0.36 | 0.83 | 0.50 | 0.59 | 0.37 | 0.39 | 0.27 | 0.60 | 0.36 | 0.47 | 0.32 | 0.57 | 0.41 | 0.66 | 0.47 | 0.36 | 0.17 | 0.51 | 0.31 | 0.47 | 0.39 | 0.37 | 0.31 | 0.48 | 0.42 | 0.44 | 0.28 | 0.44 | 0.36 | 0.45 | 0.25 | 0.31 | 0.20 | 0.44 | 0.29 |
| **$\mu$-GBP-DropConnect** | 0.71 | 0.29 | 0.85 | 0.36 | 0.47 | 0.29 | 0.50 | 0.29 | 0.64 | 0.25 | 0.36 | 0.51 | 0.69 | 0.25 | 0.70 | 0.29 | 0.41 | 0.33 | 0.52 | 0.24 | 0.21 | 0.29 | 0.22 | 0.25 | 0.23 | 0.27 | 0.21 | 0.23 | 0.21 | 0.25 | 0.27 | 0.34 | 0.16 | 0.19 | 0.16 | 0.23 |
| **$\sigma$-GBP-DropConnect** | 0.69 | 0.30 | 0.85 | 0.42 | 0.45 | 0.29 | 0.45 | 0.30 | 0.63 | 0.29 | 0.33 | 0.44 | 0.69 | 0.25 | 0.67 | 0.28 | 0.37 | 0.30 | 0.49 | 0.23 | 0.19 | 0.28 | 0.23 | 0.25 | 0.23 | 0.27 | 0.21 | 0.25 | 0.23 | 0.23 | 0.27 | 0.36 | 0.16 | 0.18 | 0.16 | 0.21 |
| **$\mu$-IG-DropConnect** | 0.72 | 0.38 | 0.85 | 0.42 | 0.50 | 0.25 | 0.46 | 0.32 | 0.62 | 0.33 | 0.46 | 0.41 | 0.63 | 0.33 | 0.64 | 0.49 | 0.43 | 0.37 | 0.63 | 0.35 | 0.28 | 0.22 | 0.30 | 0.24 | 0.31 | 0.26 | 0.32 | 0.24 | 0.26 | 0.23 | 0.39 | 0.28 | 0.27 | 0.24 | 0.31 | 0.20 |
| **$\sigma$-IG-DropConnect** | 0.72 | 0.34 | 0.83 | 0.46 | 0.43 | 0.25 | 0.38 | 0.32 | 0.65 | 0.36 | 0.45 | 0.30 | 0.62 | 0.31 | 0.65 | 0.47 | 0.42 | 0.31 | 0.60 | 0.33 | 0.29 | 0.25 | 0.29 | 0.23 | 0.30 | 0.26 | 0.34 | 0.24 | 0.26 | 0.22 | 0.37 | 0.28 | 0.28 | 0.26 | 0.26 | 0.21 |
| **$\mu$-GBP-Flipout** | 0.70 | 0.27 | 0.92 | 0.39 | 0.59 | 0.38 | 0.51 | 0.33 | 0.51 | 0.22 | 0.45 | 0.39 | 0.59 | 0.20 | 0.73 | 0.35 | 0.31 | 0.23 | 0.55 | 0.25 | 0.27 | 0.32 | 0.31 | 0.29 | 0.32 | 0.33 | 0.46 | 0.27 | 0.27 | 0.25 | 0.32 | 0.24 | 0.39 | 0.28 | 0.28 | 0.19 |
| **$\sigma$-GBP-Flipout** | 0.61 | 0.26 | 0.92 | 0.47 | 0.57 | 0.40 | 0.44 | 0.34 | 0.51 | 0.23 | 0.41 | 0.32 | 0.59 | 0.20 | 0.69 | 0.33 | 0.29 | 0.22 | 0.54 | 0.22 | 0.30 | 0.31 | 0.30 | 0.30 | 0.32 | 0.35 | 0.45 | 0.27 | 0.26 | 0.26 | 0.32 | 0.25 | 0.36 | 0.24 | 0.29 | 0.21 |
| **$\mu$-IG-Flipout** | 0.70 | 0.36 | 0.89 | 0.49 | 0.61 | 0.42 | 0.48 | 0.36 | 0.58 | 0.29 | 0.48 | 0.26 | 0.52 | 0.22 | 0.67 | 0.44 | 0.41 | 0.21 | 0.64 | 0.29 | 0.30 | 0.31 | 0.35 | 0.33 | 0.40 | 0.36 | 0.48 | 0.33 | 0.30 | 0.26 | 0.44 | 0.22 | 0.39 | 0.33 | 0.29 | 0.32 |
| **$\sigma$-IG-Flipout** | 0.69 | 0.35 | 0.88 | 0.52 | 0.60 | 0.42 | 0.43 | 0.40 | 0.56 | 0.29 | 0.49 | 0.24 | 0.56 | 0.20 | 0.66 | 0.44 | 0.38 | 0.22 | 0.64 | 0.29 | 0.38 | 0.32 | 0.35 | 0.36 | 0.40 | 0.39 | 0.48 | 0.35 | 0.29 | 0.29 | 0.47 | 0.23 | 0.39 | 0.32 | 0.31 | 0.32 |

Table 5. Class-wise AUCs for the pixel flipping plots for CIFAR-10 (columns 1-10) and FER+ (columns 11-18). The **+** and the **-** denote the pixel insertion and pixel deletion metrics respectively. The combination of the uncertainty estimation methods (Deep Ensemble, MC-Dropout, MC-DropConnect and Flipout, explanation methods (Guided BackPropagation (**GBP**) and Integrated Gradients (**IG**)), and type of explanation heatmap (mean ($\mu$) and standard deviation ($\sigma$)) are listed along the rows.

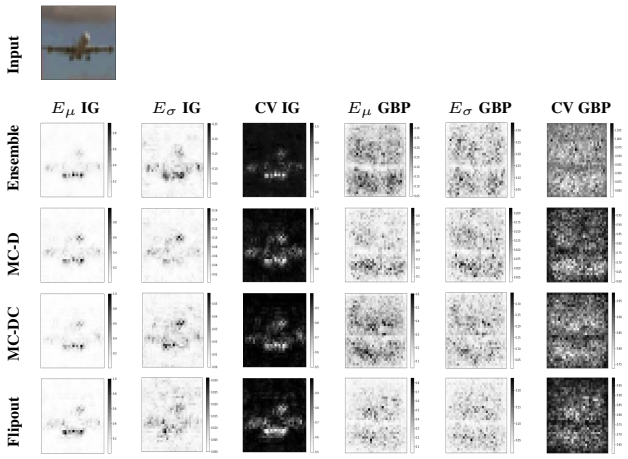


Figure 11. Example explanations of an airplane from CIFAR-10.

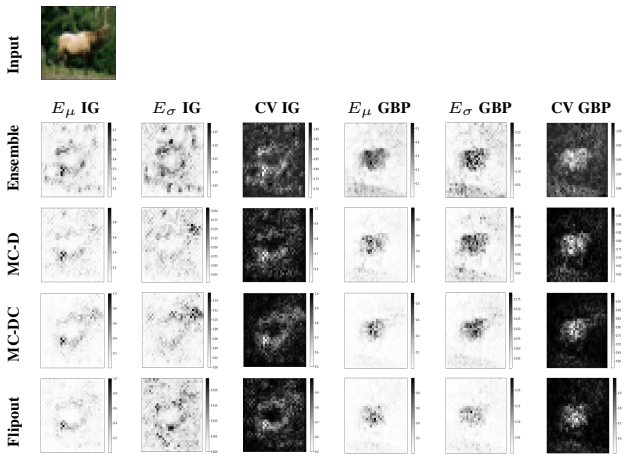


Figure 12. Example explanations of a deer from CIFAR-10.

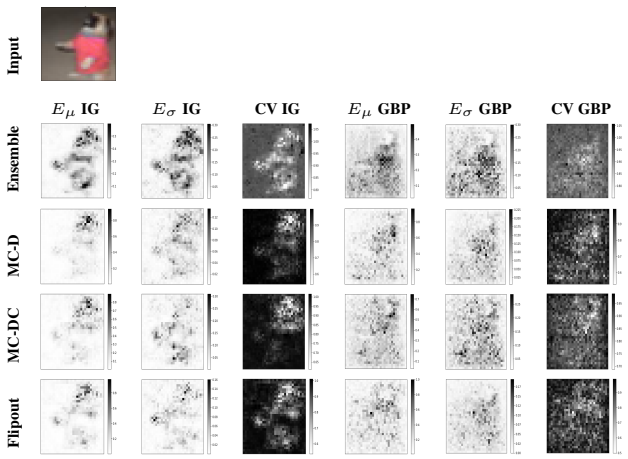


Figure 13. Example explanations of a dog taken from CIFAR-10.

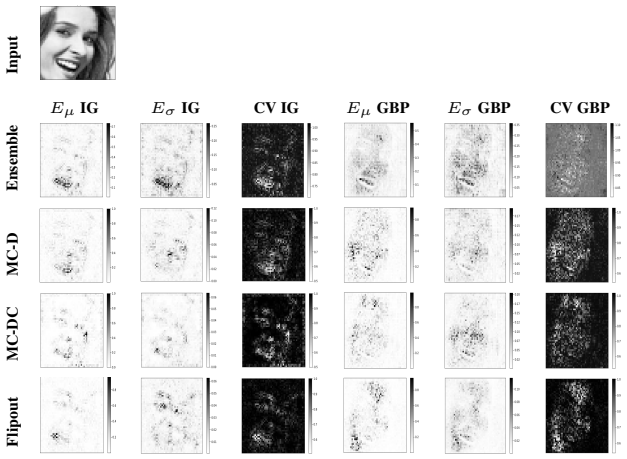


Figure 14. Example explanations of a happy person from FER+.

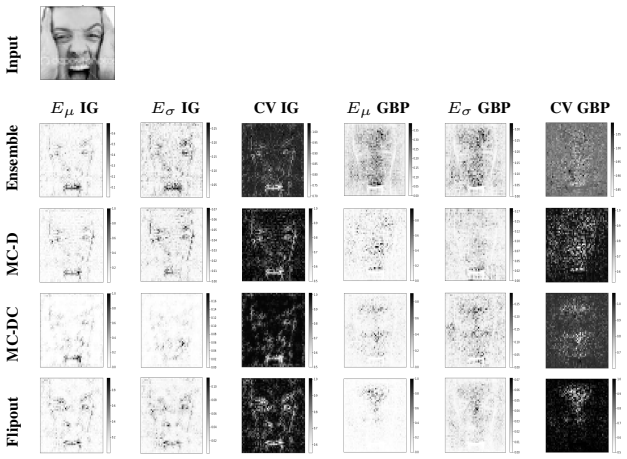


Figure 15. Example explanations of an angry person from FER+.

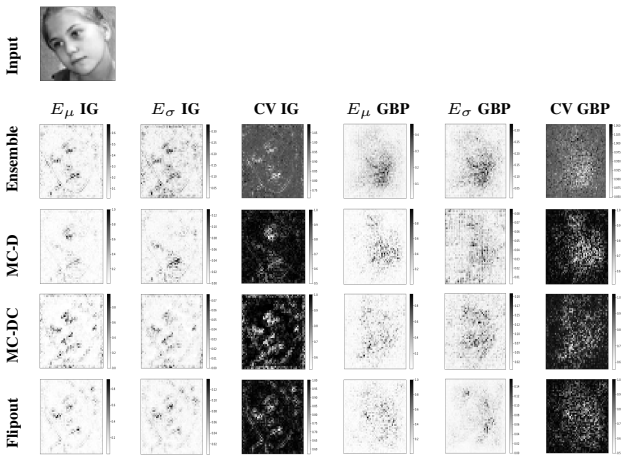


Figure 16. Example explanations of a neutral person from FER+.