1. SUPPLEMENTAR MATERIAL

The tables presented in this section provide a detailed analysis of the methods and metrics used to evaluate facial features and eeriness in different characters. They include information on the identification of the most relevant facial regions, the performance of feature extraction techniques, and the evaluation of metrics related to eeriness and comfort. The results are summarized below to facilitate understanding of the main aspects evaluated.

Table 1 shows the evaluation of the three main facial parts relevant to LIME (Top1, Top2, Top3) to identify strangeness for each character. The Agreement column indicates the accordance between the region considered most strange according to LIME (ROI) and the ground truth (GT) with 38.09% of the correct class. A "-" in the column indicates the absence of strangeness, according to the GT.

Table 2 shows the results of the Voting Regressor (VR) technique, with the evaluation of the median values of the RMSE metric considering the extraction of features from the entire face or specific parts, different standardization methods (standard, logarithmic and normalized) and the application or not of dimensionality reduction (PCA). The average elapsed time is presented in minutes.

Table 3 shows the evaluation of characters with the extraction of complete and specific facial features (forehead, eyes, nose, mouth, chin) by AUs, Entropy, GLCM, Golden Ratio, and Hu Moments. The columns PFF1, PFF2, and PFF3 show the percentage of frames attributed to the predominant feature, while MVF1, MVF2, and MVF3 indicate which feature had the highest PFF value.

Due to space constraints, we focus on an in-depth discussion of only four selected characters (from 40) in this section, where three are uncomfortable, and one is comfortable.However, it is important to note that we achieved 38.09% accuracy in predicting face parts compared to the ground truth (see Table 1), highlighting the challenging nature of this topic for study. In this section, we show the results obtained by the VR model to predict the CCS comfort and also explain this prediction in terms of the most uncomfortable part of the face. It is important to note that we achieved 61.90% accuracy in prediction face parts compared to the ground truth (see Table **??**).

Table 4 shows that when evaluating the first relevant variable (Top1) identified by LIME as the part of the face that causes strangeness, the accuracy with GT2 is 19.04%. When we evaluate only the second variable (Top2), we have 23.80% accuracy. If we evaluate only the third variable (Top3), there is an increase in accuracy of 28.57%. Therefore, when evaluating the first 3 most relevant variables identified by LIME, we obtain an accuracy of 61.90%.

Figure 1 shows feature interpretability by analyzing the training data (excluding character 8) and the test data (video frames of character 8). On the left, feature importance is highlighted, with the forehead and chin as key discomfort areas. On the right, LIME analysis for frame 125 reveals that mouth asymmetry, eye shape, and nose shape positively influence predictions, while chin shape and forehead elongation contribute negatively. The results confirm face discomfort predictions, emphasizing the mouth, eyes, and nose as primary areas of concern.

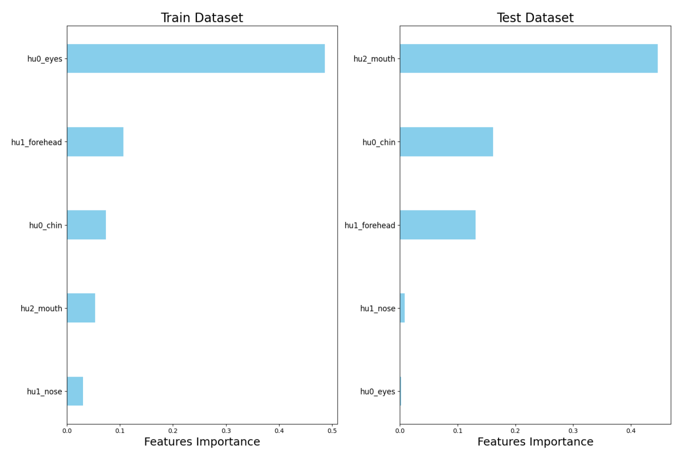

Fig. 1. Global analysis of the relevance of the features in the training (left) and testing (right) datasets for character 8. The figure on the right, corresponding to the testing dataset for character 8, covering all frames, highlights a predominance of importance in the features of the mouth and chin, agreeing with the evaluations in GT Face only in relation to the prominence of the chin, since the part of the face selected by the participants is the chin as being uncomfortable.

Figure 3 illustrates the VR model's feature interpretability, comparing the training dataset (left) and the testing dataset with video frames of character 9 (right). Figure 4 highlights the LIME analysis for frame 69 of character 9, showing positive contributions (orange) from the forehead, eyes, and nose shapes, and negative contributions (blue) from the mouth asymmetry and chin shape. These insights confirm the VR

character	ROI	GT	Agreements with ROI			
			Top1	Top2	Top3	Agreement
1	eyes	1	Agree	Disagree	Disagree	Agree
2	-	0	-	-	-	-
3	eyes	1	Agree	Disagree	Disagree	Agree
4	eyes	1	Agree	Disagree	Agree	Agree
5	mouth	1	Disagree	Disagree	Disagree	Disagree
6	-	0	-	-	-	-
7	mouth	1	Disagree	Disagree	Disagree	Disagree
8	chin	1	Disagree	Disagree	Disagree	Disagree
9	-	0	-	-	-	-
10	forehead	1	Disagree	Agree	Agree	Agree
11	-	0	-	-	-	-
12	-	0	-	-	-	-
13	eyes	1	Disagree	Disagree	Disagree	Disagree
14	-	0	-	-	-	-
15	mouth	1	Disagree	Disagree	Disagree	Disagree
16	-	0	-	-	-	-
17	mouth	1	Disagree	Disagree	Disagree	Disagree
18	-	0	-	-	-	-
19	-	0	-	-	-	-
20	eyes	1	Agree	Disagree	Disagree	Agree
21	eyes	1	Disagree	Disagree	Agree	Agree
22	-	0	-	-	-	-
23	-	0	-	-	-	-
24	eyes	1	Disagree	Disagree	Disagree	Disagree
25	eyes	1	Disagree	Disagree	Disagree	Disagree
26	mouth	1	Disagree	Disagree	Disagree	Disagree
27	-	0	-	-	-	-
28	mouth	1	Disagree	Disagree	Disagree	Disagree
29	eyes	1	Disagree	Disagree	Agree	Agree
30	-	0	-	-	-	-
31	mouth	1	Disagree	Disagree	Disagree	Disagree
32	-	0	-	-	-	-
33	-	0	-	-	-	-
34	mouth	1	Disagree	Disagree	Disagree	Disagree
35	forehead	1	Agree	Agree	Disagree	Agree
36	-	0	-	-	-	-
37	mouth	1	Disagree	Disagree	Disagree	Disagree
38	-	0	-	-	-	-
39	-	0	-	-	-	-
40	-	0	-	-	-	-

Table 1. Evaluation of the first 3 features (Top1, Top2, Top3) relevant to LIME as causing strangeness. The evaluation is performed for each feature. The Agreement column is the evaluation made considering the 3 features. If one of them agrees with the ROI column, which is the ground truth of the part of the face considered strangest, then the result is Agree, otherwise it is Disagree. The ”-” in the Agreement column indicates that according to the GT column (ground truth of the entire face) they do not generate strangeness. There was agreement of 38.09% of the characters in relation to the parts of the GT face.

Feature	Face	Standard	Logarithm	Normalized	PCA	Median RMSE	Median Elapsed Time
AUs	All Face	n	n	y	n	18.55%	79.26
					y	**17.57%**	105.31
			y	n	n	19.76%	60.02
					y	20.11%	94.99
		y	n	n	n	20.29%	59.91
					y	21.16%	95.27
Entropy	All Face	n	n	y	n	**13.88%**	83.00
					y	19.98%	69.33
			y	n	n	17.52%	103.96
					y	21.97%	54.70
		y	n	n	n	21.01%	104.17
					y	21.59%	85.31
Golden Ratio	All Face	n	n	y	n	18.74%	145.82
					y	19.42%	61.15
			y	n	n	22.22%	85.50
					y	**18.12%**	38.74
		y	n	n	n	20.50%	107.81
					y	22.14%	74.07
GLCM	All Face	n	n	y	n	23.08%	141.83
					y	21.52%	30.42
			y	n	n	**16.79%**	107.16
					y	18.09%	19.78
		y	n	n	n	19.05%	106.10
					y	23.79%	48.48
Hu Moments	All Face	n	n	y	n	**16.44%**	43.07
					y	19.28%	16.96
			y	n	n	17.59%	38.93
					y	24.41%	21.41
		y	n	n	n	23.64%	35.30
					y	26.64%	20.74
	Face Parts	n	n	y	n	16.80%	54.95
					y	20.89%	26.23
			y	n	n	**15.55%**	61.15
					y	22.22%	32.59
		y	n	n	n	20.93%	62.79
					y	25.40%	38.52

Table 2. Results of VR technique. Evaluation of the median values of the RMSE metric according to the extraction of features from the whole face or parts of it, standardization of the data (standard, logarithmic and normalized) and whether dimensionality reduction (PCA) is performed or not. The Median Elapsed Time column is shown in minutes.

Character	MVF1	PFF1	MVF2	PFF2	MVF3	PFF3
1	Hu Moments	100.00%	Hu Moments (parts)	7.14%	Golden Ratio	22.69%
2	Hu Moments	6.67%	Hu Moments	45.00%	AU	86.67%
3	Entropy	100.00%	Hu Moments (parts)	74.79%	Hu Moments	4.20%
4	Entropy	56.06%	Hu Moments (parts)	80.30%	Golden Ratio	78.79%
5	Hu Moments	3.78%	Hu Moments	58.40%	Entropy	82.79%
6	Entropy	93.53%	GLCM	56.03%	Golden Ratio	66.81%
7	GLCM, Hu Moments (parts)	100.00%	Entropy	42.31%	-	0.00%
8	GLCM	100.00%	AU	81.51%	Golden Ratio	9.24%
9	Entropy	5.13%	Entropy	82.91%	Hu Moments	93.16%
10	Hu Moments (parts)	71.85%	Hu Moments	86.13%	AU	68.07%
11	-	0.00%	-	0.00%	Hu Moments (parts)	100.00%
12	-	0.00%	GLCM	83.13%	Hu Moments (parts)	100.00%
13	Hu Moments (parts)	98.31%	Entropy	27.12%	Entropy, Hu Moments	3.39%
14	-	0.00%	Golden Ratio, GLCM	5.98%	Hu Moments (parts)	96.58%
15	Hu Moments (parts)	77.97%	GLCM	59.32%	Hu Moments	18.64%
16	AU	0.78%	GLCM	68.39%	Golden Ratio	50.52%
17	-	0.00%	Hu Moments (parts)	3.70%	Hu Moments	26.39%
18	Hu Moments	100.00%	Hu Moments (parts)	66.67%	Entropy	3.33%
19	Entropy	99.76%	GLCM	84.29%	Hu Moments (parts)	99.76%
20	GLCM	98.92%	Hu Moments (parts)	87.10%	Entropy	3.23%
21	Entropy	81.44%	AU	52.49%	Hu Moments	41.53%
22	Hu Moments (parts)	72.28%	Entropy	71.46%	Hu Moments	21.36%
23	AU	23.41%	AU	68.29%	Hu Moments (parts)	85.37%
24	GLCM	17.68%	Entropy	36.59%	Entropy, Hu Moments	54.88%
25	Hu Moments	97.90%	Entropy	69.99%	AU	11.34%
26	Hu Moments	3.09%	Hu Moments	25.26%	Hu Moments	45.36%
27	AU, GLCM	100.00%	Hu Moments (parts)	78.79%	Golden Ratio	3.03%
28	-	0.00%	Hu Moments	20.75%	Entropy	28.63%
29	GLCM	96.88%	AU	100.00%	Entropy	6.25%
30	Hu Moments (parts)	64.66%	Hu Moments	84.55%	AU	67.80%
31	Entropy	47.66%	AU	50.84%	AU, Golden Ratio	31.93%
32	AU	76.73%	Golden Ratio	44.49%	Entropy	33.88%
33	AU	65.22%	Golden Ratio	45.85%	Hu Moments	51.38%
34	AU	92.45%	GLCM	32.08%	Hu Moments (parts)	26.42%
35	Golden Ratio	82.77%	Hu Moments	40.76%	Entropy	14.14%
36	Entropy	75.68%	Hu Moments (parts)	86.49%	Golden Ratio	38.74%
37	Hu Moments	95.02%	AU	66.92%	AU	30.35%
38	AU, GLCM, Hu Moments, Hu Moments (parts)	100.00%	Golden Ratio	15.46%	Golden Ratio	3.09%
39	-	0.00%	Hu Moments (parts)	7.14%	Entropy	100.00%
40	-	0.00%	Hu Moments	9.09%	Hu Moments	88.43%

Table 3. Characters whose full faces and specific parts (forehead, eyes, nose, mouth, chin) had features extracted by AUs, Entropy, GLCM, Golden Ratio, Hu Moments and Hu Moments (parts) and evaluated with the RMSE metric. Columns PFF1, PFF2 and PFF3 show the percentage of frames of each character attributed to the feature with the highest percentage. Columns MVF1, MVF2 and MVF3 indicate which of the six features achieved the highest PFF value.

character	ROI	GT	Agreements with ROI			
			Top1	Top2	Top3	Agreement
1	eyes	1	Disagree	Disagree	Agree	Agree
2	-	0	-	-	-	-
3	eyes	1	Disagree	Agree	Disagree	Agree
4	eyes	1	Disagree	Disagree	Disagree	Disagree
5	mouth	1	Disagree	Agree	Disagree	Agree
6	-	0	-	-	-	-
7	mouth	1	Disagree	Disagree	Disagree	Disagree
8	chin	1	Disagree	Disagree	Disagree	Disagree
9	-	0	-	-	-	-
10	forehead	1	Disagree	Agree	Disagree	Agree
11	-	0	-	-	-	-
12	-	0	-	-	-	-
13	eyes	1	Agree	Disagree	Disagree	Agree
14	-	0	-	-	-	-
15	mouth	1	Disagree	Disagree	Disagree	Disagree
16	-	0	-	-	-	-
17	mouth	1	Disagree	Disagree	Agree	Agree
18	-	0	-	-	-	-
19	-	0	-	-	-	-
20	eyes	1	Agree	Disagree	Disagree	Agree
21	eyes	1	Disagree	Agree	Agree	Agree
22	-	0	-	-	-	-
23	-	0	-	-	-	-
24	eyes	1	Disagree	Agree	Agree	Agree
25	eyes	1	Agree	Disagree	Disagree	Agree
26	mouth	1	Disagree	Disagree	Agree	Agree
27	-	0	-	-	-	-
28	mouth	1	Disagree	Disagree	Disagree	Disagree
29	eyes	1	Disagree	Disagree	Disagree	Disagree
30	-	0	-	-	-	-
31	mouth	1	Disagree	Disagree	Disagree	Disagree
32	-	0	-	-	-	-
33	-	0	-	-	-	-
34	mouth	1	Disagree	Disagree	Agree	Agree
35	forehead	1	Agree	Disagree	Disagree	Agree
36	-	0	-	-	-	-
37	mouth	1	Disagree	Disagree	Disagree	Disagree
38	-	0	-	-	-	-
39	-	0	-	-	-	-
40	-	0	-	-	-	-

Table 4. Evaluation of the first 3 features (Top1, Top2, Top3) relevant to LIME as causing strangeness. The evaluation is performed for each feature. The Agreement column is the evaluation made considering the 3 features. If one of them agrees with the ROI column, which is the ground truth (GT2) of the part of the face considered strangest, then the result is Agree, otherwise it is Disagree. The ”-” in the Agreements column indicates that according to the GT column (ground truth of the entire face) they do not generate strangeness.

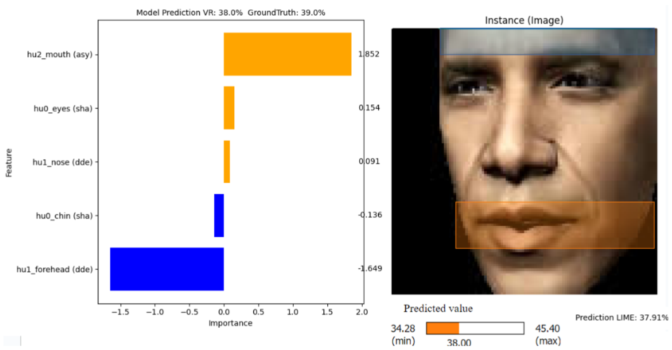

Fig. 2. Interpretability by LIME for character 8 on frame 125. On the left it shows the probability of the classes, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

and LIME model's comfortable prediction for frame 69, consistent with the CCS and GT Face.

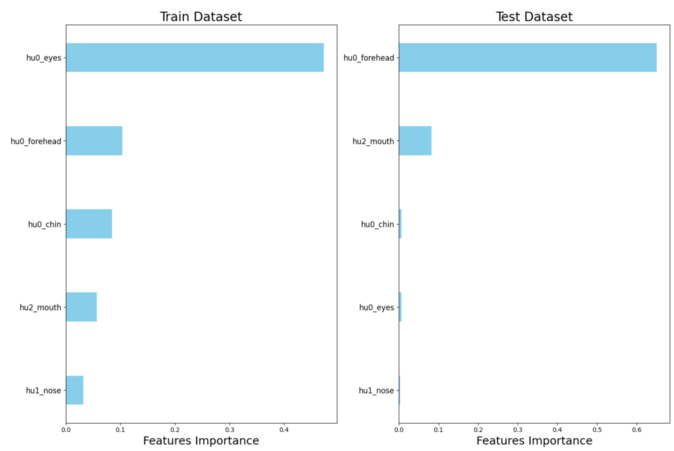

Fig. 3. Global analysis of feature relevance in training (left) and testing (right) datasets for character 9. The figure on the right, corresponding to the testing dataset for character 9, covering all frames.

Figure 5 shows feature interpretability for character 26's training and testing datasets. Figure 6 highlights interpretability using LIME at frame 34, predicting discomfort, consistent with GT Face. Key areas like the nose and mouth align with GT as relevant, while LIME also identifies potential comfort-inducing features (e.g., eyes, forehead, chin). This raises future questions about whether specific facial features can still cause discomfort even in generally comfortable characters, though this was not explored in the current study.

The table 3 shows the values of the SPF3F column, which represents the sum of the frame percentages (RMSE) for the first three tracks of each character. The Predominant Feature column identifies the relevant algorithms. The conditions for calculating the SPF3F allow us to identify whether

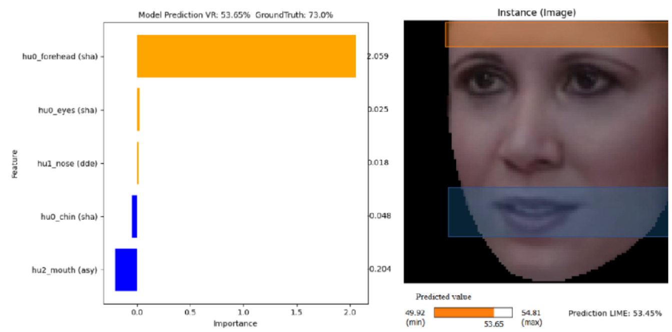

Fig. 4. Interpretability by LIME for character 9 in table 69. On the left shows the comfort prediction of the face (CCS), in the middle the weights generated by the model for each relevant feature and on the right the evaluated face.

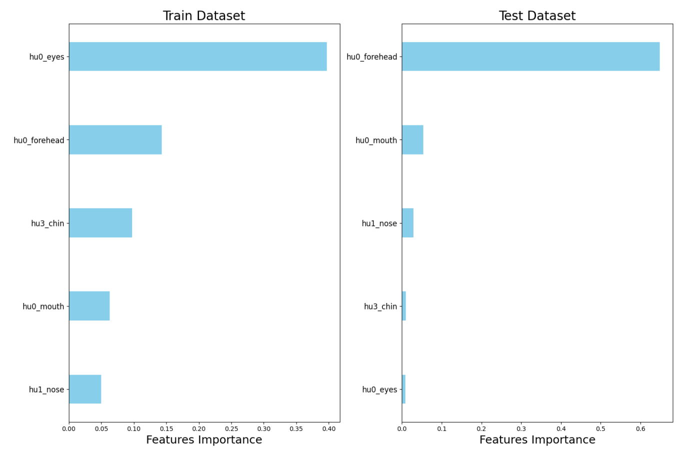

Fig. 5. Global analysis of the relevance of features in the training (left) and testing (right) datasets of character 26. The figure on the right, corresponding to the testing dataset of character 26, covering all frames. It highlights a predominance of importance in the features of the forehead, nose and mouth. The part of the face selected by the participants is the mouth, as shown .

one or more features were dominant or shared in each character. It can be seen that for 15 characters, the Predominant Feature column indicates the predominance of the Hu Moments and AU features, which appear separately in several cases. We can mention character 1 (Hu Moments, 100.00%), character 2 (AU, 87.00%), and characters 31, 32 and 34 (AU, with 51.00%, 77.00% and 92.00%, respectively). The Entropy feature also presents a case of relevant predominance in character 39, with 100.00%. In 25 characters, the Predominant Feature column displays a combination of two or more features. This indicates the need to consider the average of the percentages according to the defined evaluation rule. Character 3, for example, presents Entropy and Hu Moments (parts), with a SPF3F of 87.50%. Character 4 has three predominant

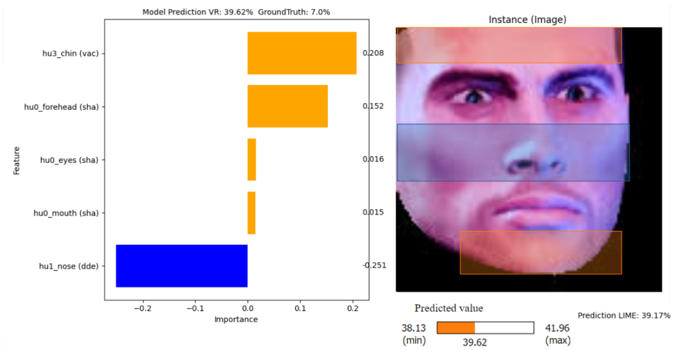

Fig. 6. Interpretability by LIME for character 26 in frame 34. On the left shows the prediction, in the middle the weights generated by the model for each relevant feature and on the right the evaluated face. The prediction agrees with GT . The characteristics that contributed positively (orange) to this result were mainly the variations in chin curvature (var), the shape of the forehead, eyes and mouth. In contrast, the nose (dde) when evaluating the direction and degree of elongation contributed negatively (blue) to this result. However, the part of the face selected by the participants is the mouth.

features: Entropy, Hu Moments (parts) and Golden Ratio, totaling 71.67%. Similar situations occur with character 10 (Hu Moments (parts), Hu Moments and AU) and character 27 (AU, GLCM and Hu Moments (parts)), with 75.33% and 89.50%, respectively. In only two characters (17 and 28), the SPF3F value is considerably low, with 15.00% and 25.00%, respectively. In these cases, there is a predominance of cumulative features, such as Hu Moments (parts) and Hu Moments in character 17, and Hu Moments and Entropy in character 28. These results suggest a greater dispersion of percentages between the analyzed ranges. By further analyzing the Predominant Features column, we identified possible frequency patterns:

- Hu Moments and Hu Moments (parts) are the most recurrent features, appearing alone or in combination in more than half of the cases. They are predominant in characters such as 1, 11, 13, 18, 26 and 30.
- AU also stands out, especially in isolated cases such as characters 2, 31, 32 and 34, in addition to appearing in combinations of features (characters 10, 23 and 27).
- Entropy appears as predominant in many combined cases, such as in characters 3, 4, 5 and 22, in addition to being unique in character 39 (100.00%).
- GLCM and Golden Ratio appear less frequently, but remain relevant when composing combinations, as observed in characters 6, 8, 16 and 35.

Characters with SPF3F values of 100.00% indicate an absolute dominance of the features listed in the Predominant Features column. We observe that this occurs in characters 1 (Hu Moments), 7 (GLCM, Hu Moments (parts)), 11 (Hu Moments (parts)), 38 (AU, GLCM, Hu Moments, Hu Moments (parts)) and 39 (Entropy). These cases represent situations in which the RMSE is completely concentrated in the first three bands for the features involved. The analysis seems to show that Hu Moments (parts) and AU are the features with the greatest influence on the cumulative percentages (SPF3F), followed by Entropy in several contexts. The predominance of cumulative features, in approximately 62.5% of the cases (25 characters), reinforces the complexity of the distribution of RMSE frames and the need to consider multiple features together. The identification of cases with SPF3F = 100.00% demonstrates situations of absolute dominance, which can be considered references for future analyses.

Character	SPF3F	Predominant Features
1	100.00%	Hu Moments
2	87.00%	AU
3	87.50%	Entropy, Hu Moments (parts)
4	71.67%	Entropy, Hu Moments (parts), Golden Ratio
5	70.50%	Hu Moments, Entropy
6	72.33%	Entropy, GLCM, Golden Ratio
7	100.00%	GLCM, Hu Moments (parts)
8	91.00%	GLCM, AU
9	88.00%	Entropy, Hu Moments
10	75.33%	Hu Moments (parts), Hu Moments, AU
11	100.00%	Hu Moments (parts)
12	91.50%	GLCM, Hu Moments (parts)
13	98.00%	Hu Moments (parts)
14	97.00%	Hu Moments (parts)
15	68.50%	Hu Moments (parts), GLCM
16	59.50%	GLCM, Golden Ratio
17	15.00%	Hu Moments (parts), Hu Moments
18	83.50%	Hu Moments, Hu Moments (parts)
19	94.67%	Entropy, GLCM, Hu Moments (parts)
20	93.00%	GLCM, Hu Moments (parts)
21	66.50%	Entropy, AU
22	71.50%	Hu Moments (parts), Entropy
23	76.50%	AU, Hu Moments (parts)
24	55.00%	Entropy, Hu Moments
25	84.00%	Hu Moments, Entropy
26	24.33%	Hu Moments, Hu Moments, Hu Moments
27	89.50%	AU, GLCM, Hu Moments (parts)
28	25.00%	Hu Moments, Entropy
29	98.50%	GLCM, AU
30	72.67%	Hu Moments (parts), Hu Moments, AU
31	51.00%	AU
32	77.00%	AU
33	58.00%	AU, Hu Moments
34	92.00%	AU
35	83.00%	Golden Ratio
36	81.00%	Entropy, Hu Moments (parts)
37	81.00%	Hu Moments, AU
38	100.00%	AU, GLCM, Hu Moments, Hu Moments (parts)
39	100.00%	Entropy
40	88.00%	Hu Moments

Table 5. Analysis of the frame percentage values RMSE) in the first three bands, with the SPF3F column representing the weighted sum of these values under three distinct conditions: (1) When only one frame percentage value exceeds 0.5, the value is directly assigned to the sum, and the corresponding algorithm is identified; (2) When more than one frame percentage value exceeds 0.5, the sum is obtained by averaging the values involved, and the cumulative algorithms are listed; (3) When all percentage values are less than or equal to 0.5, the same logic of averaging and cumulative indication of the algorithms is applied. The Predominant Algorithm column identifies the algorithms that participate in the rule applied in the SPF3F column.