
MIX: A Multi-view Time-Frequency Interactive Explanation Framework for Time Series Classification

Anonymous Author(s)

Affiliation

Address

email

1 Supplementary Material

2 In this Supplementary Material, we provide an in-depth analysis of our proposed attribution mechanisms, IGV and KIGV, in Section A. Next, we present additional experimental results and comprehensive ablation studies in Section B. Finally, in Section C, we showcase qualitative visualizations that illustrate the interpretability of our method across different time-frequency views.

6 The source code and datasets used in our experiments are included in the ZIP file.

7 Contents

8	A Analysis on Keystone-first Integrated Gradients	1
9	B Extensive Experimental Results	2
10	B.1 Extensive Comparison	2
11	B.2 Extensive Ablation Studies	2
12	C Qualitative Results	3
13	C.1 Visualization of Explanations on UCR Dataset	3
14	C.2 Visualization of Explanations on Synthetic Dataset	4
15	C.3 Visualization of Explanations on MIT-BIH Dataset	4

16 A Analysis on Keystone-first Integrated Gradients

17 In both the main paper and Appendices, we have shown that IGV, satisfies the axioms of Integrated
18 Gradients (IG) and our overall attribution mechanism is implementation invariant. Here, we further
19 analyze KIGV within the cross-view refinement framework to demonstrate its advantages over
20 standard IG. By prioritizing the explanation of "keystone" features, KIGV effectively reduces the
21 noise present in importance scores generated by IG, resulting in more accurate attribution maps.
22 As a result, OSIGV also benefits from these improved scores. As illustrated in Fig. 1, our method
23 avoids highlighting irrelevant features, which is further supported by the higher $KAUC\hat{S}_{top}$ values
24 achieved with KIGV compared to IG.

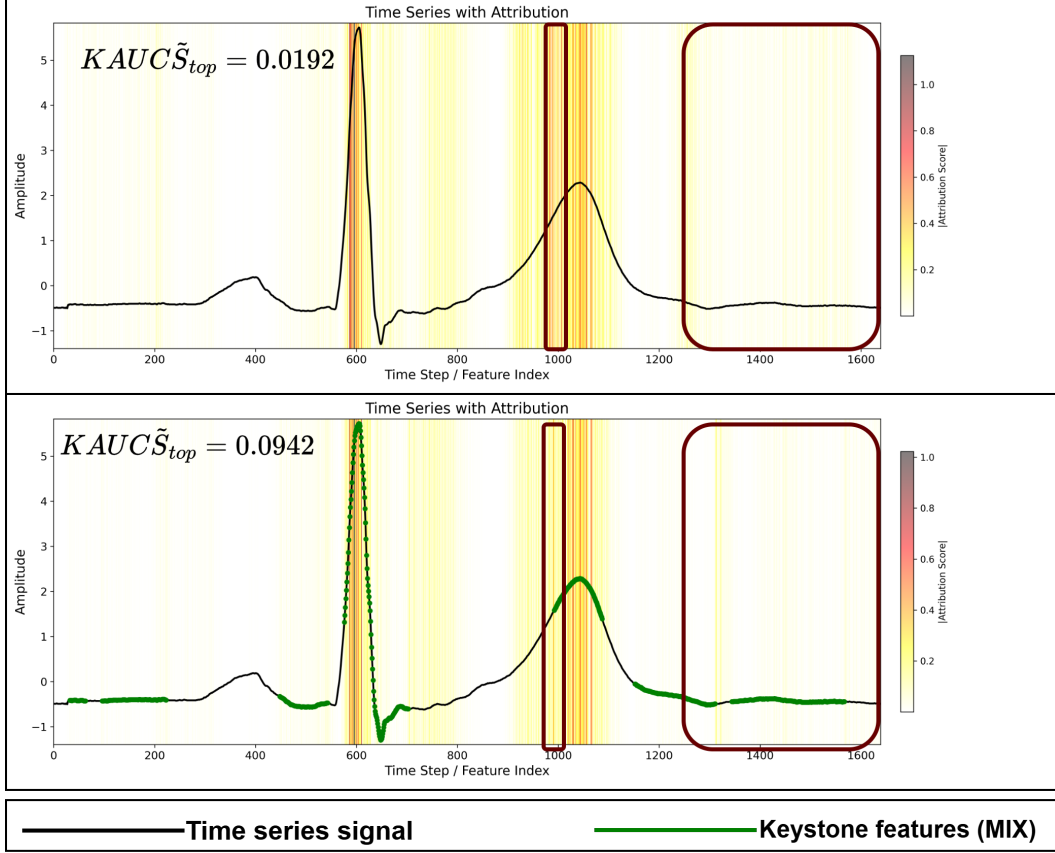


Figure 1: Comparison of KIGV (bottom) with conventional Integrated Gradients (IG, top) on the CinCEGTorso dataset. The time series signal is shown as a black line, while the highlighted segment in green indicates the keystone parts used as input for KIGV. The background color represents the attribution value at each time step. In the regions highlighted by red rectangles, KIGV avoids attributing importance to less relevant features that are mistakenly emphasized by standard IG. This demonstrates the effectiveness of cross-view refinement in filtering out irrelevant features. It is also reflected by higher $KAUC\tilde{S}_{top}$ with KIGV.

25 B Extensive Experimental Results

26 B.1 Extensive Comparison

27 In the main paper and Appendices, we primarily compare our method to other approaches at the
 28 best-performing wavelet level. However, it is also intriguing to study the effectiveness of cross-view
 29 refinement for cA_0 . To assess the quality of explanations at wavelet level 0 (i.e., the raw time series),
 30 we compare MIX with state-of-the-art methods specifically designed for this domain, including
 31 LIMESegment and InteDisUX. As shown in Fig. 2, our method consistently outperforms both
 32 baselines in terms of faithfulness and robustness.

33 B.2 Extensive Ablation Studies

34 Along with ablation studies on overall Phase 2, the attribution mechanism (compared to standard
 35 IG), and the effects of window and step size, we also investigate the adaptive selector, which is a
 36 novel component in cross-view refinement. The adaptive selector dynamically determines whether to
 37 adopt the newly generated explanation or retain the previous one, thereby maintaining the quality
 38 of the explanations. Our ablation results, shown in Fig. 3, demonstrate that the adaptive selector
 39 consistently enhances explanation quality across all wavelet levels on both evaluated datasets.

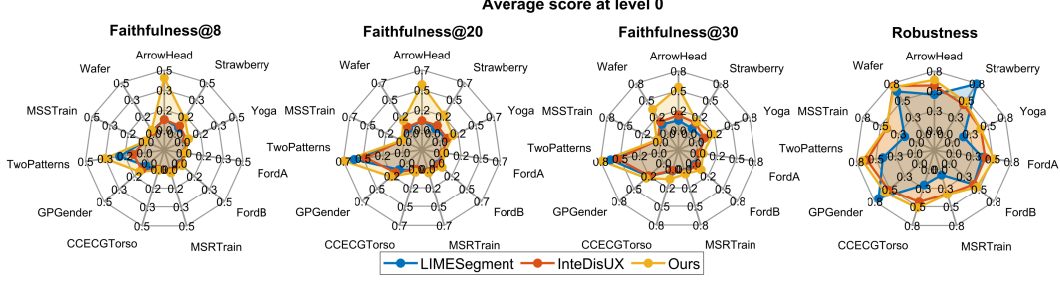


Figure 2: Comparison of explanations generated by our method (MIX) at wavelet level 0 in terms of faithfulness@8, faithfulness@20, faithfulness@30, and robustness for each dataset average over 3 DL architectures. Our proposed approach consistently outperforms state-of-the-art methods, LIMEsegment and InteDisUX, on all metrics.

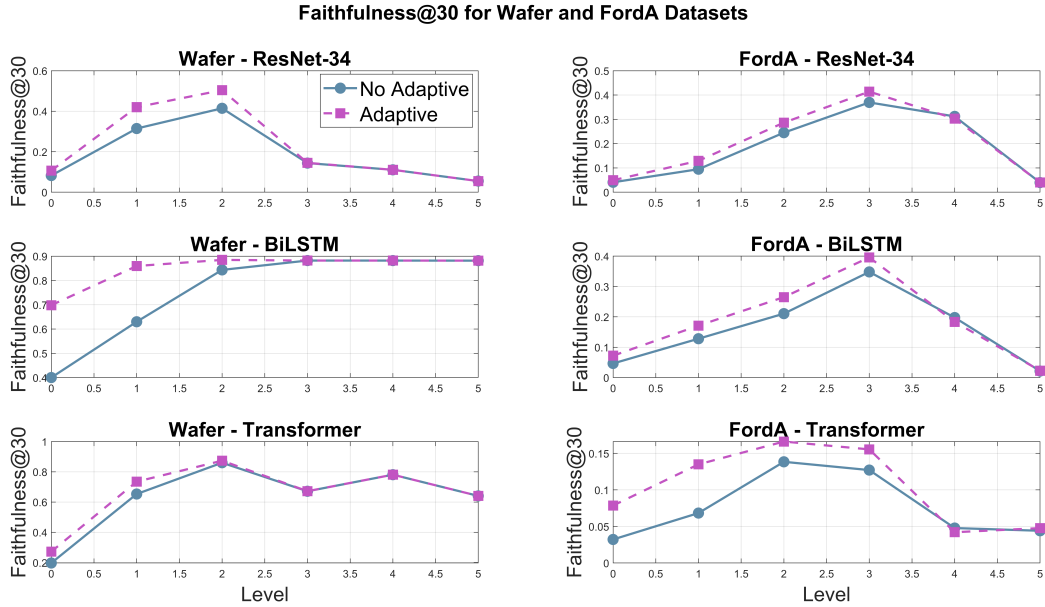


Figure 3: Ablation study on the adaptive selector for the Wafer and FordA datasets across three deep learning architectures. The results show that incorporating the adaptive selector (adaptive line) improves explanation quality across all wavelet levels compared to using KIGV alone (no adaptive line).

40 C Qualitative Results

41 C.1 Visualization of Explanations on UCR Dataset

42 **CinCECGTorso dataset.** We present a visualization of our explanations on the CinCECGTorso
 43 dataset and ResNet-34 DL model across multiple instances for each class in Fig. 4. The figure
 44 suggests that explanations at wavelet level 1 or 2 tend to capture relevant features more effectively
 45 than those at wavelet level 0. Furthermore, the multi-view visualization for a single instance in Fig. 5
 46 indicates that explanation quality is highest at level 3, not at cA_0 . Taken together, these visualizations
 47 highlight the potential value of a multi-view setup for explanation compared to relying solely on the
 48 raw time series.

49 **Wafer dataset.** We present a visualization of our explanations on the Wafer dataset using a ResNet-34
 50 deep learning model to illustrate how explanations can improve when generated from different views
 51 (see Fig. 6). These results suggest that employing a multi-view setup in the wavelet domain may
 52 enhance the interpretability of deep learning models for time series classification. Furthermore, we

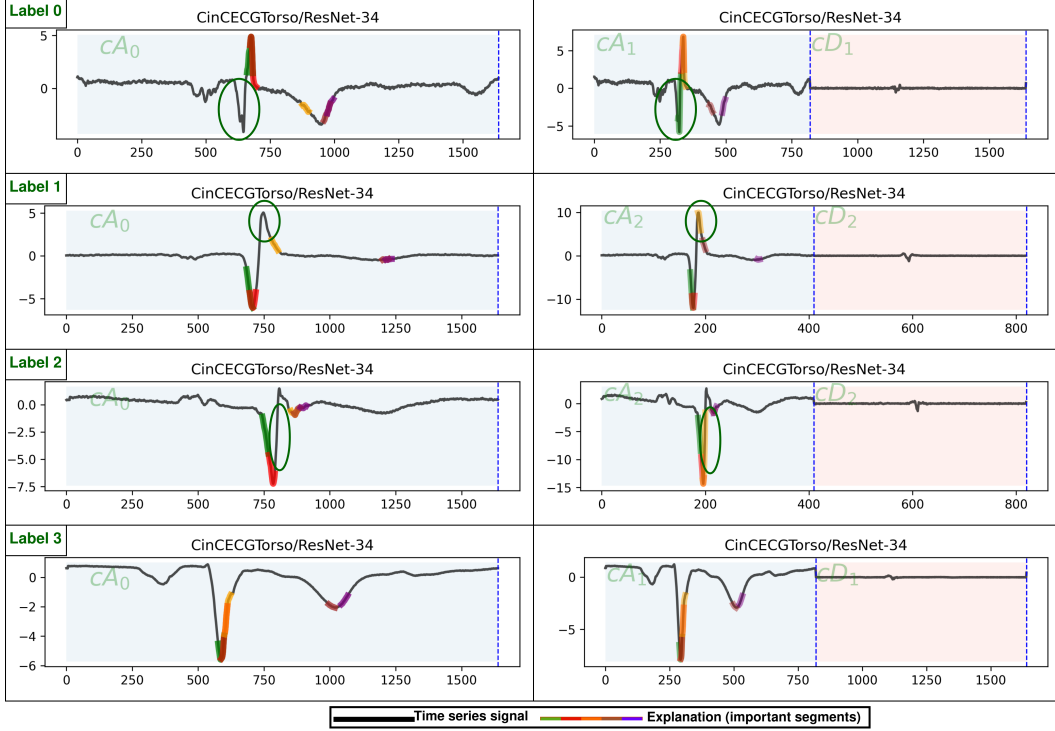


Figure 4: Visualization of the **CinCECGTorso** time series signal (black line) at wavelet level 0 (cA_0) and the best wavelet level for explanation for each instance (ranging from level 0 to 5), along with the top explanatory segments (colorful lines, each assigned a distinct color such as green, red, orange, brown, or purple) generated by MIX. Each row represents one instance. The background regions shaded in blue and pink correspond to cA and cD components, respectively, with their names indicated inside each region. The four signals correspond to four different class labels, and differences among these labels can be observed in the visualizations. Key features that are correctly captured by MIX at wavelet levels 1 or 2 (cA_1 , cA_2) but missed at level 0 (raw time series) are indicated with green circles. These results highlight the importance of incorporating multiple DWT views for explanation, demonstrating that a multi-view approach can provide more comprehensive and accurate explanations than relying on the raw time series alone.

53 observe that each view can both miss and capture important features that are not detected by the other,
 54 indicating that cooperation between views is beneficial to provide better explanations.

55 **TwoPatterns dataset.** We present a visualization of our explanations on the TwoPatterns dataset and
 56 ResNet-34 DL model in Fig. 7 to compare explanations at level 0 versus level 1, and in Fig. 8 for
 57 level 2. The results indicate that MIX produces better explanations at levels 1 compared to level 0
 58 and level 2 and level 0 can support each other. This finding further underscores the importance of a
 59 multi-view setup, consistent with our observations on the CinCECGTorso and Wafer datasets.

60 C.2 Visualization of Explanations on Synthetic Dataset

61 We also provide a visualization of explanations on the synthetic dataset. As shown in Fig. 9 and
 62 Fig. 10, MIX effectively captures important features in cD_1 , which aligns with the way the data was
 63 synthesized.

64 C.3 Visualization of Explanations on MIT-BIH Dataset

65 For real-world applications, we visualize our explanations on the MIT-BIH dataset. As shown
 66 in Fig. 11, MIX captures important features more effectively by utilizing wavelet levels 2 and 3,
 67 rather than relying solely on the raw time series. This leads to a natural question: is it sufficient to
 68 search for a single optimal explanation, or is a more flexible, greedy strategy across multiple views

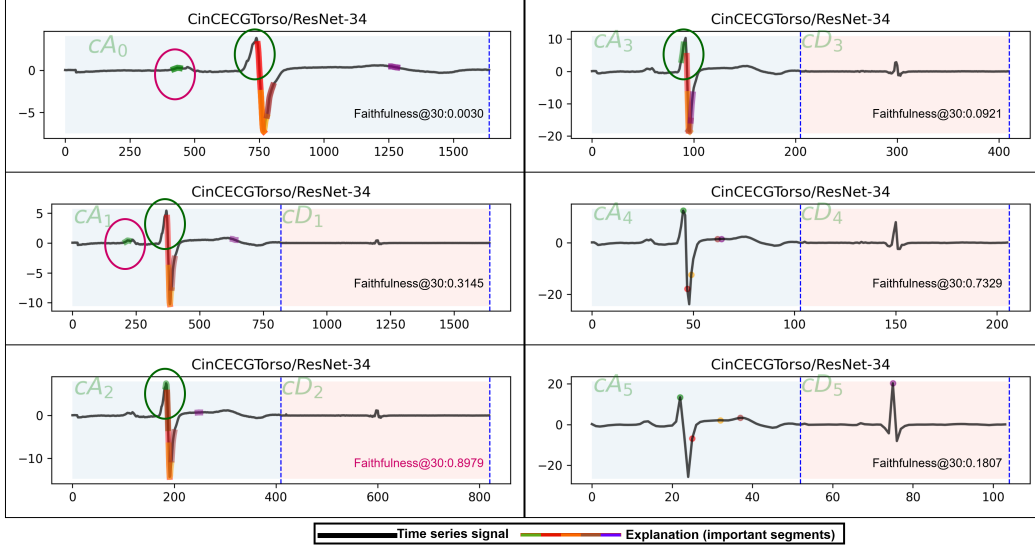


Figure 5: Visualization of the **CinCECGTorso** time series signal (black line) at wavelet level 0 (cA_0) and higher levels (cA_1 through cA_5), together with their corresponding explanations generated by MIX. All figures correspond to a single instance across five wavelet levels. For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. The black line represents the time series signal for each component, while the colorful lines indicate the top segments identified by MIX, with each segment assigned a distinct color, including green, red, orange, brown, and purple. Key regions that are missed at levels 0 and 1 but captured at higher levels are marked with green circles. Notably, some important segments are successfully identified in cA_2 and cA_3 but not in cA_0 , which is consistent with the Faithfulness@30 evaluation (highest at level 3 and lowest at level 0) shown in each subfigure. At higher levels (cA_4 , cA_5), explanations may also miss crucial segments and are generally less effective than those at cA_3 . Additionally, cA_0 sometimes captures noise (highlighted with a pink circle), which is eliminated at level 3 and above. These results further highlight the importance of a multi-view explanation setup.

69 preferable? Fig. 12 provides further insight—depending on the instance, explanations at cA_3 may
70 capture important features those are missed at cA_4 , and vice versa. These observations reinforce
71 the importance of a multi-view setup and highlight the value of a greedy selection strategy that
72 traverses multiple DWT levels to produce more faithful and robust explanations. Notably, this finding
73 is consistent with our observations on the Wafer and TwoPatterns datasets.

74 References

- 75 [1] Owen Queen, Tom Hartvigsen, Teddy Koker, Huan He, Theodoros Tsiligkaridis, and Marinka
76 Zitnik. Encoding time-series explanations through self-supervised model behavior consistency.
77 *NeurIPS*, 36, 2024.

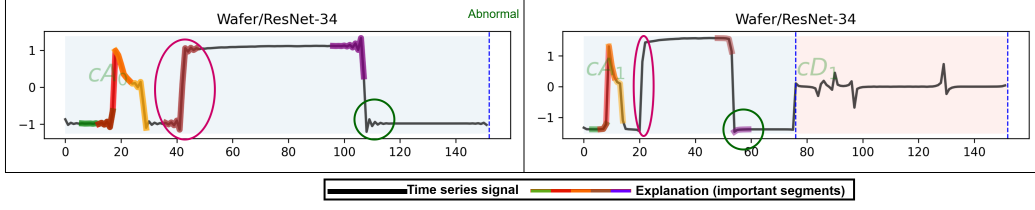


Figure 6: Visualization of the **Wafer** time series signal (black line) with top segments from MIX's explanation represented as colorful lines, each color corresponding to a different top segment, for the two classes: Normal and Abnormal. Each row corresponds to one instance. For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. At level 0, MIX can capture features associated with the Abnormal class; however, a small segment (highlighted with a green circle) is missed at this level but is successfully identified in cA_1 (right figure). Conversely, some important segments captured in cA_0 (highlighted with a pink circle) are not identified at cA_1 . These results indicate that different views can complement each other to achieve more faithful explanations, supporting the effectiveness of our multi-view setup for time series explanation.

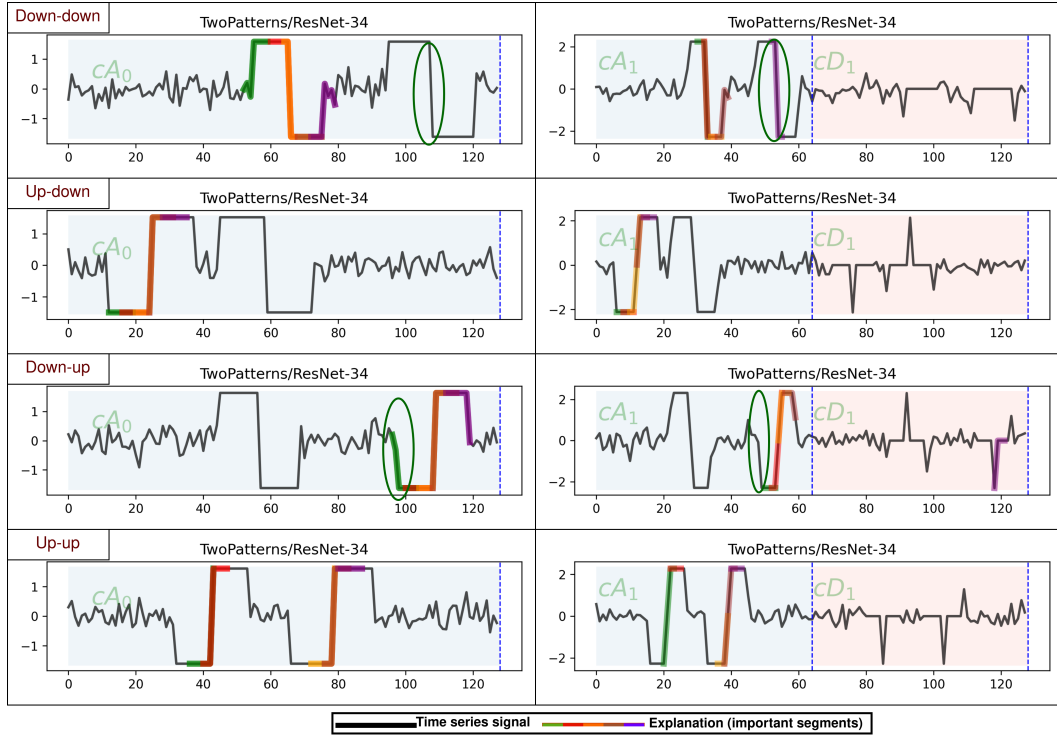


Figure 7: Visualization of the **TwoPatterns** time series signal (black line) with top segments from MIX's explanation represented as colorful lines, each color corresponding to a different top segment, for the four classes represented by two patterns: down-down, up-down, down-up, and up-up. Each row corresponds to one instance across two wavelet levels, 0 and 1. For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. For the "down-down" class, MIX fails to capture the segment corresponding to the 2nd "down" pattern in cA_0 , but the explanation in cA_1 successfully highlights this segment as important. Similarly, for the "down-up" class, the segment for the "down" pattern has a redundant part recognized as important segment (marked in green circle) in cA_0 , but it is not captured in cA_1 . These findings indicate the importance of a multi-view setup for time series classification (TSC) explanation.

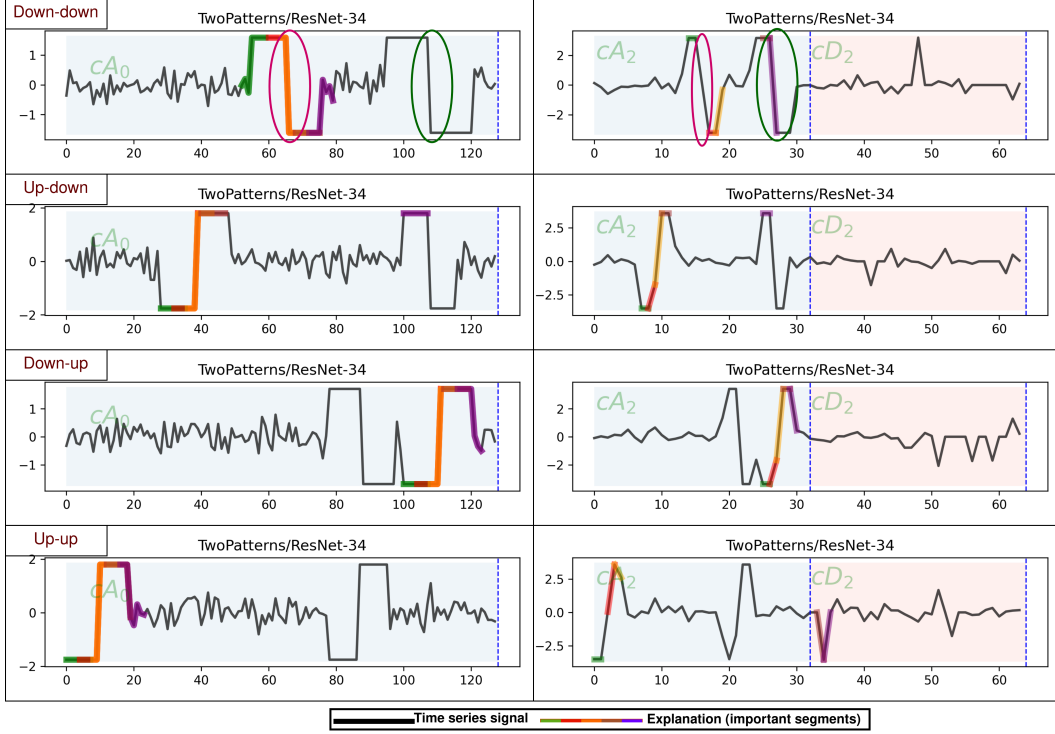


Figure 8: Visualization of the **TwoPatterns** time series signal (black line) with top segments from MIX's explanation represented as colorful lines, each color corresponding to a different top segment, for the four classes defined by two patterns: down-down, up-down, down-up, and up-up. Each row corresponds to one instance across two wavelet levels (0 and 2). For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. For the "down-down" class, MIX fails to capture the segment corresponding to the "down" pattern in cA_0 , but the explanation in cA_2 successfully highlights this segment as important. Conversely, an important segment highlighted by a pink circle is recognized by MIX in cA_0 but not in cA_2 . These observations indicate that each view can reveal different important features and can support each other, supporting our motivation for a multi-view explanation setup.

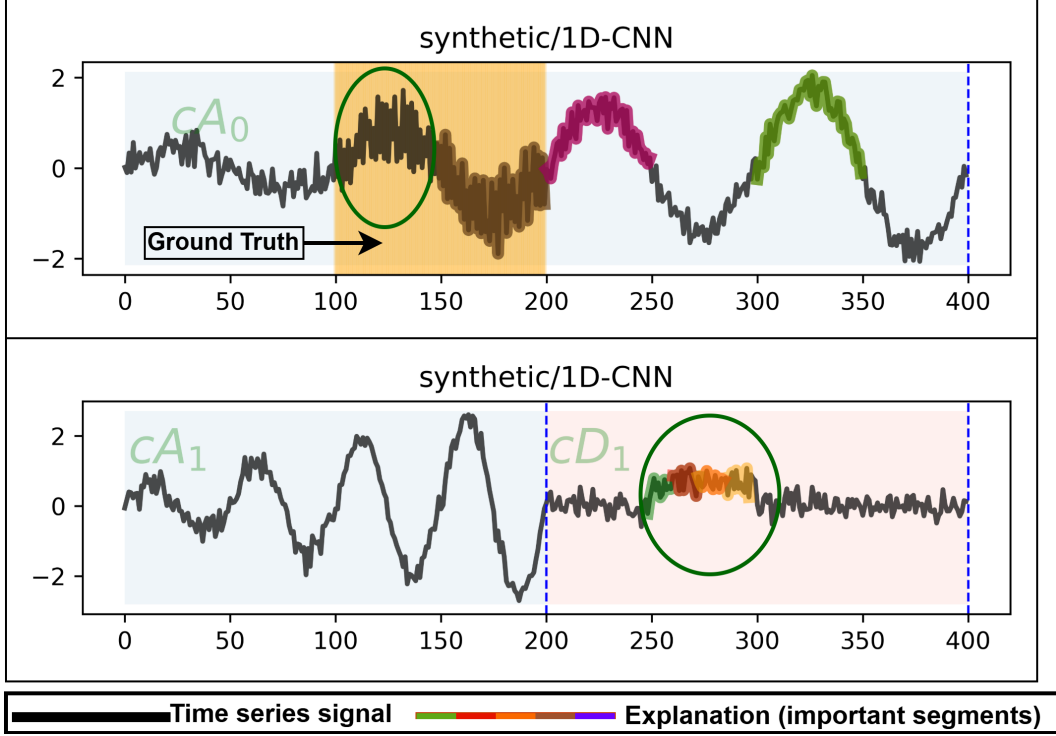


Figure 9: Visualization on the **synthetic** dataset with the top 100 important features highlighted by selecting top segments. All figures corresponds to one instance across two wavelet levels (0 and 1). For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. The ground truth region for this dataset is from time step 100 to 200 in cA_0 , corresponding to steps 250 to 300 in cD_1 based on the synthetic process, highlighted by color yellow. MIX successfully captures the key features in cD_1 , correctly identifying the important region marked in green circle (positions 250–300 in cD_1 , which maps to 100–200 in cA_0 and matches the ground truth). In contrast, the explanation on cA_0 alone misses part of the ground truth (notably, the segment from approximately time step 100 to 150 remains unhighlighted) marked in green circle, demonstrating the benefit of multi-view analysis.

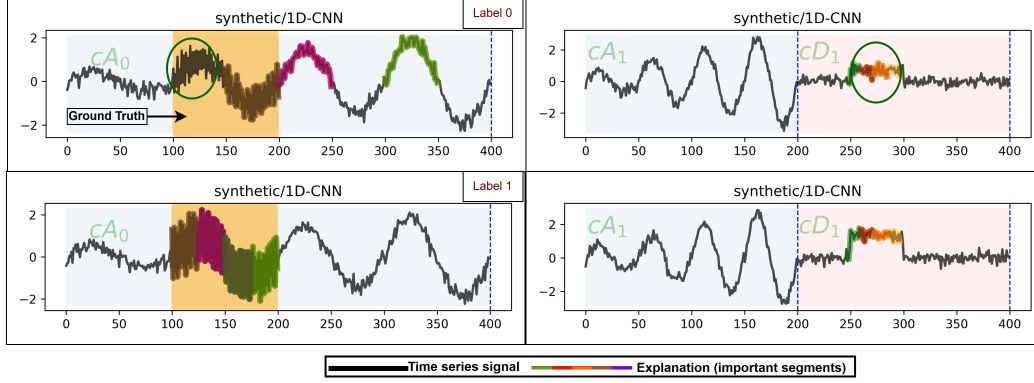


Figure 10: Visualization on the **synthetic** dataset with the top 100 important features highlighted by selecting top segments. Each row corresponds to one instance across two wavelet levels (0 and 1). For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. The ground truth region for this dataset is from time step 100 to 200 in cA_0 , corresponding to steps 250 to 300 in cD_1 based on the synthetic process, highlighted by color yellow. MIX successfully captures the key features in cD_1 , correctly identifying the important region marked with a green circle in both classes (positions 50 to 100 (250–300 as shift from cA_1) in cD_1 , mapping to 100–200 in cA_0 , matching the ground truth). In contrast, the explanation on cA_0 alone misses part of the ground truth in class 0 (notably, the segment from approximately time steps 100 to 150 remains unhighlighted, as marked by the green circle), indicating the benefit of multi-view analysis.

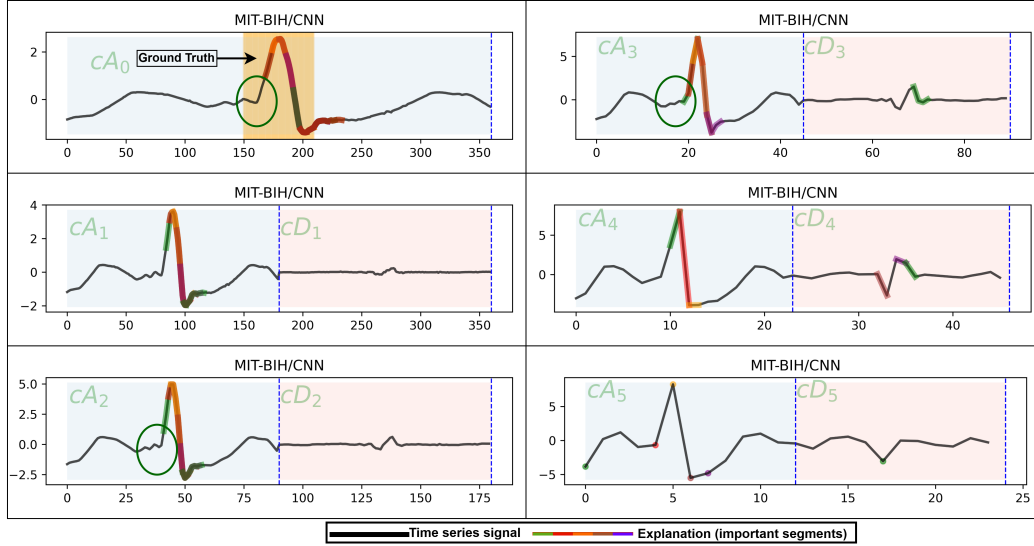


Figure 11: Visualization on the **MIT-BIH** dataset with the top segments highlighted. All figures represent a single instance across five wavelet levels. The ground truth, as annotated by cardiologists according to [1], typically corresponds to the peak region in the middle of the signal, highlighted by color yellow. For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. Explanations using only cA_0 miss some important features, whereas cA_2 , and cA_3 more effectively capture the relevant segments, which are marked with green circles.

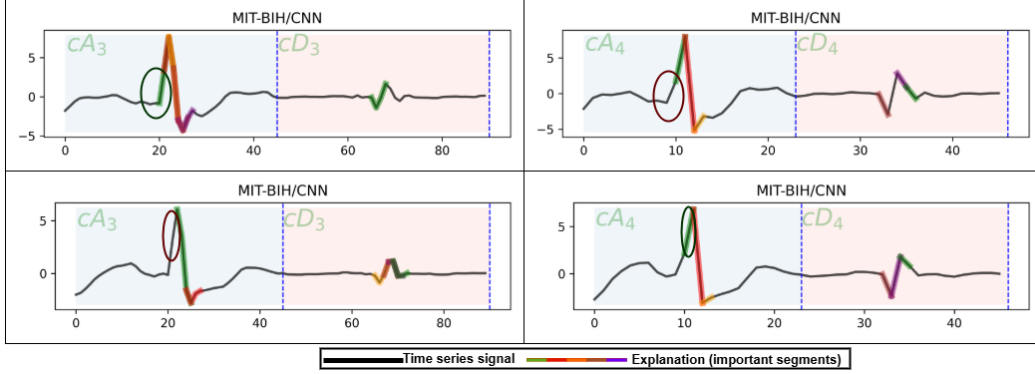


Figure 12: Visualization on the **MIT-BIH** dataset with the top segments highlighted. Each row shows one instance at two DWT levels (3 and 4). For each level l , cA_l and cD_l are shown as two regions with light blue and pink backgrounds, labeled accordingly. In the instance in the first row, the explanation in cA_3 at level 3 covers important features more effectively than cA_4 at level 4. The region captured by cA_3 is indicated with a green circle, while the corresponding region in cA_4 (marked with a red circle) is not highlighted. Conversely, for the instance in the second row, MIX at level 4 with cA_4 highlights important features that are missed at level 3 with cA_3 (with missed regions marked by red circles in cA_3 and by green circles in cA_4 where they are correctly identified). These observations suggest that a multi-view setup and cross-view aggregation are beneficial for achieving more faithful explanations.