NANYANG TECHNOLOGICAL UNIVERSITY

SC4001: Neural Network & Deep Learning

Group Project: (A)

Deep Learning for ECG Heartbeat Classification

Koh Kai Jie, Farrel (U2120219J)

Leo Zhi Kai (U2120050L)

Low Cheng Feng (U2120689E)

Contents

Abstract

The automated classification of electrocardiogram (ECG) signals using deep learning techniques represents a significant advancement in cardiac diagnostic and monitoring. This report explores various deep learning architectures for the classification of ECG heartbeats, comparing the effectiveness of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid approaches. Using the publicly available datasets from PhysioNet, we evaluated these architecture's performance in distinguishing between normal heartbeats and various types of arrhythmias. Our methodology encompasses data augmentation techniques such as noise injection and time warping to enhance the model robustness and implementing explainable AI methods such as LIME (Local Interpretable Model-agnostic Explanations) for feature importance to provide insights into the features driving classification decisions. The results demonstrate the comparative advantages of different architectural approaches with our CNN model achieving the best performance in terms of accuracy and generalization.This research contributes to the growing body of work on automated cardiac diagnosis systems, offering potential applications in both clinical settings and remote patient monitoring. These findings suggest that deep learning-based ECG classifications systems can serve as valuable tools for healthcare professionals in the rapid and accurate detection of cardiac abnormalities.

Introduction

Electrocardiograms (ECGs) are critical diagnostic tools used to monitor the electrical activity of the heart. This helps medical professionals detect various cardiac conditions in their patients, from arrhythmias to more severe heart diseases. Timely detection of conditions can help medical professionals make life-saving decisions in the delivery of appropriate treatments, which in turn improves patient outcomes. In the field of medical signal processing, deep learning methods can automate this detection process and classify heartbeats with high accuracy and efficiency. In this report we explore several methods such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) and compare the accuracies of various architectures.

Dataset

Source: https://physionet.org/content/mitdb/1.0.0/
https://www.kaggle.com/datasets/shayanfazeli/heartbeat/data

The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979. Twenty-three recordings were chosen at random from a set of 4000 24-hour ambulatory ECG recordings collected from a mixed population of inpatients (about 60%) and outpatients (about 40%) at Boston's Beth Israel Hospital; the remaining 25 recordings were selected from the same set to include less common but clinically significant arrhythmias that would not be well-represented in a small random sample.

The Arrhythmia Dataset Information:

- Number of Samples: 109446
- Number of CAtegories : 5
- Sampling Frequency: 125Hz
- Data Source: Physionet's MIT-BIH Arrhythmia Dataset
- Classes: [Normal: 0 S:1 Ventricular:2 Fibrillation:3 Q(Unclassified):4]

Each row represents an example in that portion of the dataset. The final element of each row denotes the class to which that example belongs.

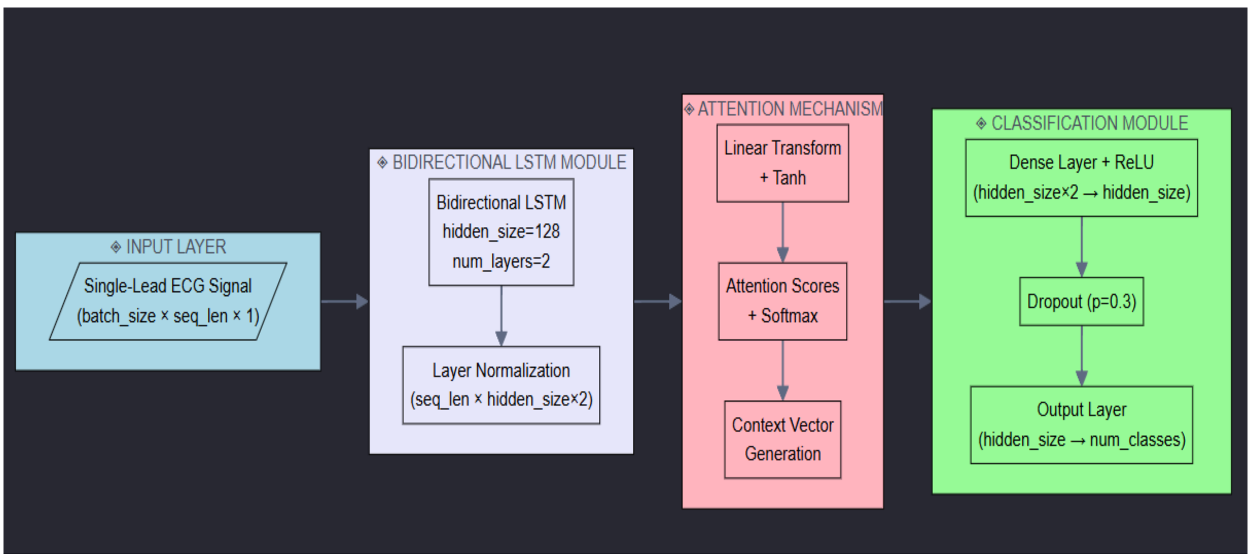
Remark: *All the samples are cropped, downsampled and padded with zeroes if necessary to the fixed dimension of 188*.

Neural Network Models

RNN Model

Our RNN Model makes use of the LSTM and Attention Neural Network Structure

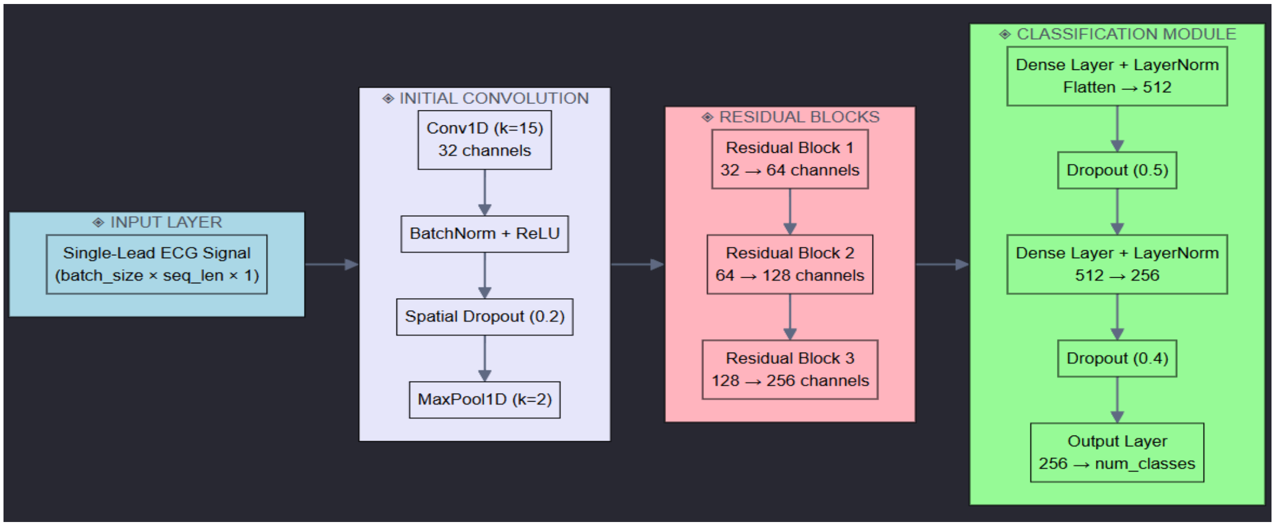
1. First the data is passed onto the LSTM forward pass,
2. Then it goes through a normalization layer
3. Next the normalized data is passed into the Attention weights
4. And Finally into a fully connected neural network with ReLU for classification.



Training Loss and Accuracy from the model over 100 Epoch
Bookmark Link:

Training Loss	Classification Result	Confusion Matrix	Test Accuracy

CNN Model

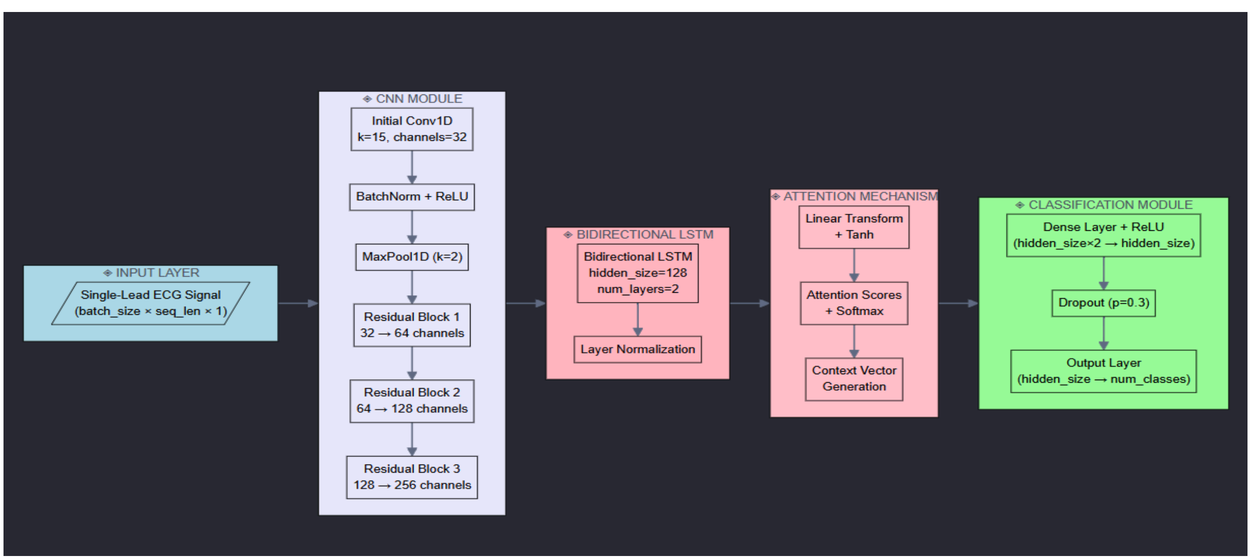


The CNN architecture we use a 1d convolutional block with a large kernel size of 15, followed by 3 improved residual blocks with skip-link connections. Finally we feed the data in 2 Fully-Connected layers. Dropout layers with varying probabilities are implemented throughout the architecture to better generalize learning. We also implemented a regularization technique called label smoothing which assigns a small probability to incorrect classes (instead of assigning 0) to prevent the model from being too confident in its predictions for a single class. This makes the model more robust to noisy labels.

Bookmark Link:

Training Loss	Classification Result	Confusion Matrix	Test Accuracy

Advanced architecture - Hybrid GRU with Attention



For our advanced neural network architecture, we used GRU along with Attention.

1. Bidirectional GRU for capturing temporal dependencies
2. Attention mechanism
3. Optional skip connection
4. Classification layers with batch normalization

Hybrid GRU with Attention trained with early stopping at patience 3 (25 epoch)

Bookmark Link:

Training Loss	Classification Result	Confusion Matrix	Test Accuracy

Experimentation

We attempted to train a hybrid model combining RNN and CNN architecture with the use of residual blocks with batch normalization. The model architecture consists of initial convolutional layers for feature extraction, followed by residual connections to mitigate the vanishing gradient problem and facilitate deeper network training. Batch normalization was implemented between convolutional layers to stabilize the learning process and reduce internal covariate shift. The network then leverages LSTM layers to capture temporal dependencies in the ECG sequences, with bidirectional processing to consider both past and future context. This hybrid approach aims to exploit both the CNN's ability to learn hierarchical spatial features from the ECG signals and the RNN's capacity to model sequential patterns, while the residual connections help maintain gradient flow through the deep architecture.

However, despite the theoretical advantages of combining both architectures, our experimental results showed that the hybrid model consistently underperformed compared to the individual CNN and RNN implementations. This unexpected performance degradation could be attributed to increased model complexity leading to optimization difficulties, or potential interference between the different architectural components during the learning process. Based on these findings, we decided to abandon the hybrid approach and focus on optimizing the individual architectural implementations separately.

We experimented on an attention-based GRU (Gated Recurrent Unit) architecture to process the ECG signals after some research into its implementation. The model incorporates a bidirectional GRU layer to capture temporal dependencies from both forward and backward directions of the sequence, enabling comprehensive feature extraction from the time series data. An attention mechanism was integrated to allow the network to dynamically focus on relevant parts of the input sequence, weighing the importance of different temporal segments during processing. This attention layer computes attention weights through a trainable mechanism that produces a context vector, helping the model identify and emphasize crucial segments of the ECG signals. The architecture also employs dropout between layers to prevent overfitting and enhance generalization. By combining GRU's efficient sequential learning capabilities with attention mechanisms, the model can better capture both local and global temporal patterns while maintaining the ability to selectively focus on diagnostically significant portions of the ECG waveforms.

The attention-based GRU architecture demonstrated superior performance in our experiments, showing strong capabilities in ECG signal processing and classification. Given these promising results, we selected this architecture as our best model for further development. To optimize the model's performance, we explored various data augmentation techniques to enhance the model's generalization capabilities and robustness.

Data Augmentation

To optimize model learning, we tested different augmentation strategies on time-series data such as time warping and noise injection, aiming to expose the model to a broader range of ECG signal variations while maintaining physiologically meaningful characteristics. The following are the various augmentation methods visualized:

1. Noise Injection

Injecting noise into your time series data can help improve the robustness of your deep learning model.

2. Time Warping

Time warping is a great technique to augment time series data by stretching or compressing the time axis. This can help improve the robustness of models, especially in tasks like classification or regression.

3. Noise and Time Wrap

Incorporating noise into time series data can enhance the robustness of deep learning models. Additionally, time warping serves as an effective technique for augmenting time series data by either stretching or compressing the temporal axis. This approach is particularly beneficial for improving model robustness in tasks such as classification and regression.

4. Details of Augmentation

The addition of ***additive noise*** significantly improved the model's performance. This type of noise, which involves adding random fluctuations to the original ECG signal, helps the model generalize better by forcing it to focus on the underlying patterns rather than memorizing specific, noise-free samples. The variability introduced by additive noise mimics real-world variations and artifacts commonly found in ECG data, enabling the model to become more robust to small perturbations and irregularities in the data. In our experiment, we improved test accuracies to 98.96% with the model trained on X values distorted by addictive noise.

We also attempted to use **multiplicative noise**. Multiplicative noise, which scales the ECG signal by a random factor, tends to distort the signal too heavily, making it harder for the model to learn meaningful features. The scaling can lead to unrealistic variations that are not representative of natural ECG patterns. This method is not suitable as ECG signals are highly structured and contain important diagnostic features in their morphology and augmenting such would negatively impact model classification performance. Thus we decided not to include multiplicative noise when augmenting our training data.

Time warping, which involves stretching or compressing the signal over time, also degraded the model's learning. Although it can, theoretically, simulate slight changes in the pacing of the heart, all the different warping values we tried have negatively impacted model accuracy. We believe it is due to the alteration of physiologically meaningful temporal relationships between the ECG signals' components and its respective classification, making it more difficult for the network to capture key features like the QRS complex or P and T waves.

In conclusion, we found out that additive noise introduced to the training data is the only type of useful data augmentation in the context of ECG dataset. Moving forward, we use additive noise to optimize model learning and achieve better classification results.

Explainability of Results

LIME Results for Class [‘Normal’] using RNN Model:

Bookmark to results: Figure 5.1

Key Observations:

- The prediction header shows "Predicted Class: Normal" with a confidence of 0.9196 (91.96%)
- The feature importance values suggest the model is looking at specific time points that are characteristic of normal ECG patterns
- The clear separation in class probabilities indicates strong discrimination between normal and abnormal patterns

This visualization suggests the model is:

1. Making a very confident correct classification
2. Using appropriate features to make its decision
3. Successfully distinguishing normal patterns from various types of abnormalities

LIME Results for Class [‘Supraventricular’] using RNN Model:

Bookmark to results: Figure 5.2

Key Observations:

- The prediction is very confident (93.74%) that this is a Supraventricular rhythm
- The feature importance graph shows which time points in the signal were most crucial for this classification
- The clear distinction in class probabilities suggests the model has learned strong discriminative features for Supraventricular patterns

Clinical Relevance:

- Supraventricular arrhythmias typically show:
    - Different timing between beats

        - Changes in P wave morphology
        - Altered QRS relationships
    - The model appears to be capturing these characteristics through its important time point features

The model is showing strong performance in:

1. Correctly identifying Supraventricular patterns
2. Using clinically relevant features for classification
3. Maintaining high confidence in its prediction

LIME Results for Class [‘Normal’] using GRU Attention RNN Model:

Bookmark to results: Figure 5.3

Key Differences from Previous Models:

- The GRU with Attention model shows:
    - Higher confidence (100%) compared to previous models
    - More distinct feature importance patterns
    - Clearer separation in class probabilities

Advantages of GRU with Attention:

1. The attention mechanism helps the model focus on the most relevant parts of the ECG signal
2. GRU units are good at capturing temporal dependencies in the signal
3. The combination seems to produce very confident and clear classifications

Notable Observations:

- The model is extremely confident in its classification (100%)
- The feature importances are well-distributed across different time points
- The attention mechanism seems to be helping identify key diagnostic features in the ECG

LIME Results for Class [Fibrillation] using GRU Attention RNN Model:

Bookmark to results: Figure 5.4

The interpretation:

- These time points likely correspond to key morphological features in the ECG signal
- The positive features (right-extending bars) indicate time points that contribute to classifying the signal into a particular class
- The negative features (left-extending bars) indicate time points that contribute against classifying the signal into that class
- The magnitude of each bar represents how strongly that feature influences the model's decision

Looking at the ECG signal:

- The most important features seem to cluster around significant waveform changes

- This suggests the model is paying attention to important clinical features like:
    - QRS complex regions
    - T-wave characteristics
    - Baseline segments

Findings: One noticeable point between RNN and GRU Attention model is the confidence of different classes.

Feature Map Explanation on CNN Model

Class 0	Class 1	Class 2	Class 3	Class 4
Figure 6.1	Figure 6.2	Figure 6.3	Figure 6.4	Figure 6.5

1. **Activation Patterns Across Classes:**
    - The feature maps exhibit distinct activation patterns for each of the 5 ECG signal classes (0, 1, 2, 3, 4).
    - This suggests the CNN model has learned to extract unique features that are relevant for discriminating between the different classes of ECG signals.
    - The diversity of activation patterns across kernels and classes indicates the model has developed a rich set of feature detectors to capture the underlying characteristics of the ECG signals.
2. **Kernel Specialization:**
    - Within each class, certain kernels (e.g., 0-53) show stronger and more varied activation patterns compared to others (54-90).
    - This implies the model has specialized some kernels to be more responsive to the distinctive features of each ECG class.
    - The suppressed or inactive kernels (54-90) may not be as relevant for the classification of the current class, but could be important for other classes.
3. **Temporal Dynamics:**
    - The temporal patterns (across the x-axis) indicate the model is capturing the dynamic nature of the ECG signals.
    - The spikes, bursts, and oscillatory patterns observed in the activations suggest the model is learning to detect temporal features and relationships within the ECG data.
    - This is particularly important for ECG classification, as the temporal characteristics of the signal carry crucial information for identifying different cardiac conditions.
4. **Feature Interaction and Complexity:**
    - The varying combinations of activation patterns, including bursts, rhythmic patterns, and localized activations, indicate the model is learning to extract complex feature representations.

- The interplay between different kernels and their activation patterns suggests the model is capturing intricate relationships and interactions within the ECG data.
- This level of complexity in the feature representations can contribute to the model's ability to accurately classify the different ECG signal classes.

5. **Interpretability and Explainability:**
   - The distinct activation patterns observed across the classes provide valuable insights into the model's decision-making process.
   - By analyzing these feature maps, you can gain a better understanding of which specific features the model is focusing on to distinguish between the different ECG classes.
   - This information can be leveraged to improve the interpretability and explainability of the CNN model, enabling clinicians and researchers to better comprehend the underlying mechanisms behind the ECG classification.

Overall, the analysis of these feature maps suggests the CNN model has developed a sophisticated set of feature detectors that are capable of capturing the complex and dynamic characteristics of the ECG signals, which is crucial for accurate classification of different cardiac conditions. The patterns observed can serve as a valuable tool for model interpretability and explainability in the context of ECG classification.

Explanation of CNN Feature Map Classes

Class 0:

- The feature maps for Class 0 show a clear distinction between the first group of kernels (0-53) and the middle group (54-90).
- Kernels 0-53 exhibit strong, varied activation patterns with a range of amplitudes and temporal characteristics, such as bursts, oscillations, and localized concentrations of activity.
- In contrast, kernels 54-90 are largely suppressed or inactive, indicating they are not as relevant for the classification of this particular class.
- The active kernels seem to be capturing diverse low-level features as well as more complex, class-specific patterns in the ECG signals.

Class 1:

- For Class 1, the feature maps demonstrate a more uniform activation across the majority of kernels, with fewer distinct groupings compared to Class 0.
- Most kernels show prominent activation patterns, suggesting the network is utilizing a broader set of feature detectors to represent the characteristics of this class.
- The temporal patterns include a mix of burst-like events, rhythmic oscillations, and concentrated regions of activity, indicating the model is learning to capture both short-term and longer-term features.
- The overall higher activation levels and more consistent patterns across kernels imply that the distinguishing features for Class 1 are more broadly distributed across the network.

Class 2:

- The Class 2 feature maps exhibit the strongest and most widespread activation patterns across all kernels, with very few suppressed or inactive regions.

- Many kernels show high-amplitude activations, ranging from around 0 to 5, suggesting the model has learned to detect highly distinctive features for this class.
- The temporal patterns are also more complex, with a mix of burst-like events, oscillatory rhythms, and localized concentrations of activity, indicating the network is capturing intricate relationships within the ECG signals.
- The consistent and diverse activation patterns across kernels imply that Class 2 has more salient and discriminative features that the model leverages for accurate classification.

Class 3:

- The feature maps for Class 3 demonstrate a pattern similar to Class 0, with a clear separation between the first group of active kernels (0-53) and the relatively suppressed middle group (54-90).
- The active kernels exhibit strong, varied activation patterns, including bursts, rhythmic oscillations, and localized regions of high activity.
- This suggests the model has learned to extract a rich set of features that are particularly relevant for the classification of Class 3, with certain kernels specializing in specific aspects of the ECG signals.
- The suppressed kernels in the middle group may be less crucial for distinguishing this class, but could play a role in the identification of other classes.

Class 4:

- The Class 4 feature maps show a more balanced activation pattern across the kernels, with fewer distinct groupings compared to some of the other classes.
- While there are still variations in the amplitude and temporal characteristics of the activations, the overall patterns appear more uniform and less specialized.
- Many kernels demonstrate prominent activation, suggesting the model is relying on a relatively broad set of features to represent the distinctive characteristics of Class 4.
- The temporal patterns include a mix of burst-like events, oscillatory rhythms, and localized concentrations of activity, indicating the network is capturing both short-term and longer-term features in the ECG signals.

Appendix

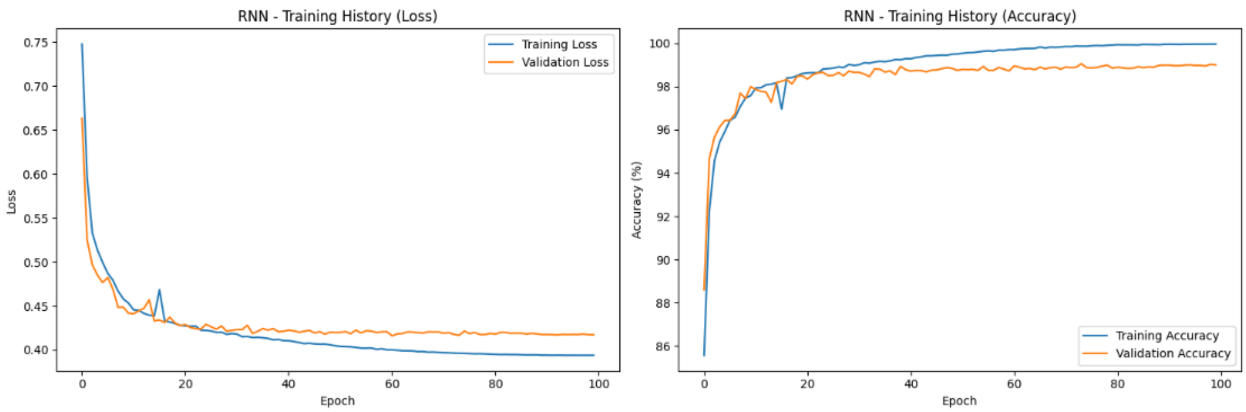


Figure 1.1 Loss and Accuracy of RNN Model

Classification Report (RNN):

	precision	recall	f1-score	support
0	0.99	1.00	0.99	18118
1	0.93	0.83	0.87	556
2	0.97	0.96	0.97	1448
3	0.86	0.84	0.85	162
4	1.00	0.99	0.99	1608
accuracy			0.99	21892
macro avg	0.95	0.92	0.94	21892
weighted avg	0.99	0.99	0.99	21892

Figure 1.2 Classification report of RNN Model

Confusion Matrix:

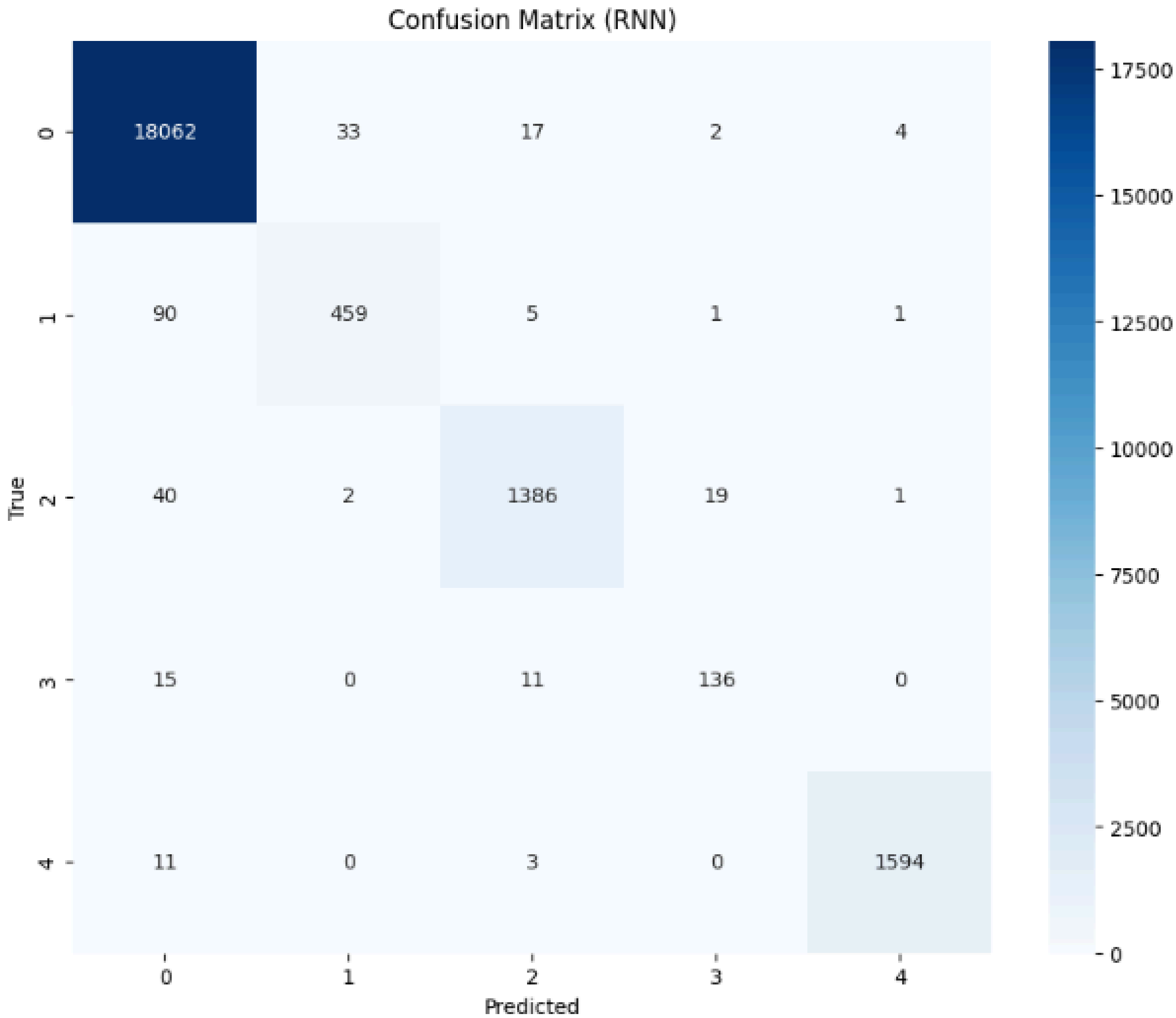


Figure 1.3 Confusion Matrix of RNN Model

```
Evaluating RNN model...
RNN Test Accuracy: 98.84%

Summary:
RNN Test Accuracy: 98.84%
```

Figure 1.4 Test Accuracy of RNN Model (98.84%)

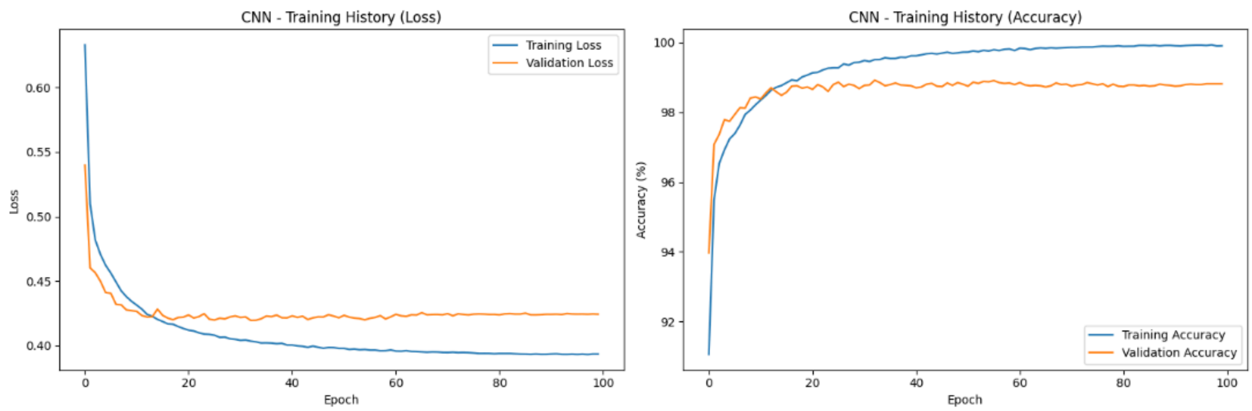


Figure 2.1 Loss and Accuracy of CNN Model

Classification Report (CNN):

	precision	recall	f1-score	support
0	0.99	1.00	0.99	18118
1	0.91	0.82	0.86	556
2	0.97	0.96	0.96	1448
3	0.86	0.77	0.81	162
4	0.99	0.99	0.99	1608
accuracy			0.99	21892
macro avg	0.95	0.91	0.93	21892
weighted avg	0.99	0.99	0.99	21892

Figure 2.2 Classification report of CNN Model

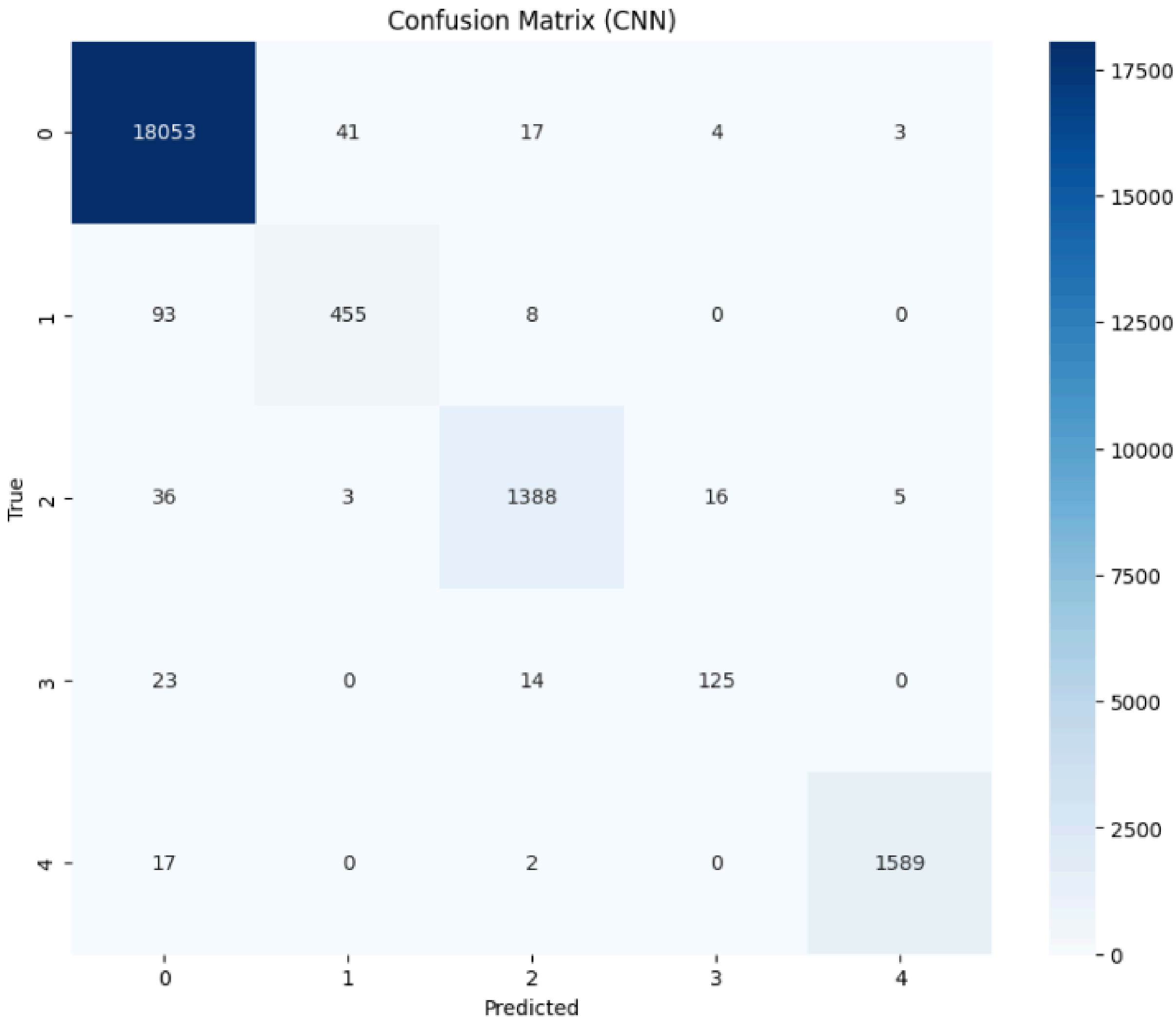


Figure 2.3 Confusion Matrix of CNN Model

```
Evaluating CNN model...
CNN Test Accuracy: 98.71%

Summary:
RNN Test Accuracy: 98.84%
CNN Test Accuracy: 98.71%
```

Figure 2.4 Test Accuracy of CNN Model (98.71%)

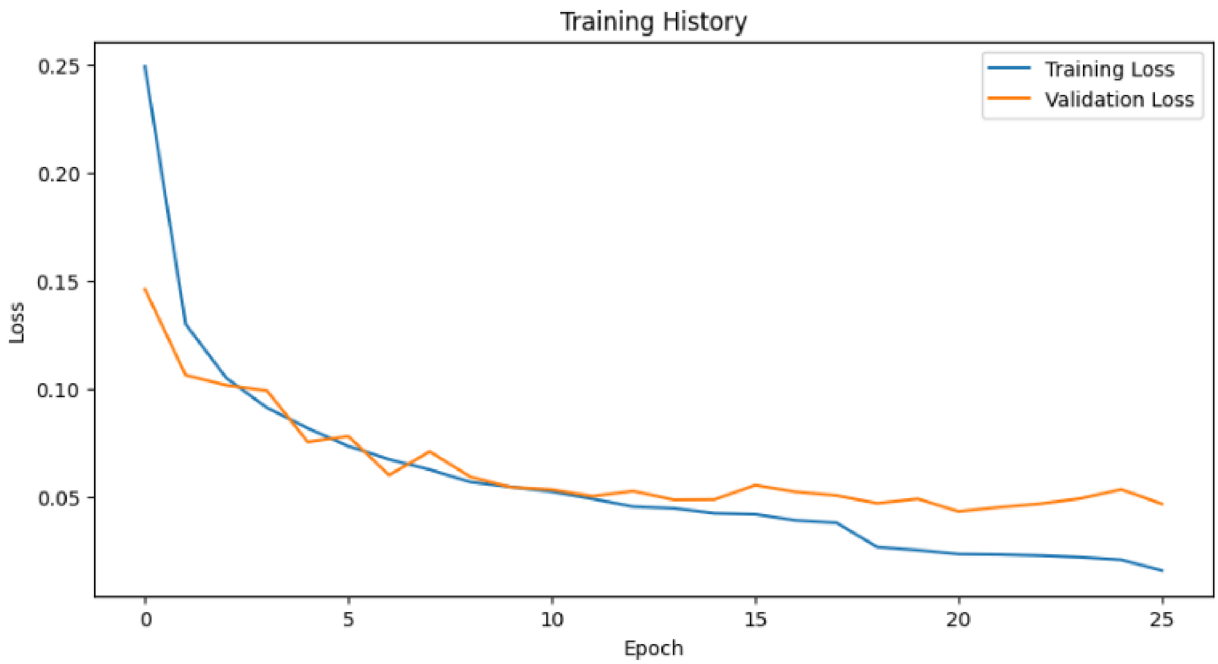


Figure 3.1 Training Loss of GRU Attention RNN Model

Classification Report:

	precision	recall	f1-score	support
0	0.99	1.00	0.99	18118
1	0.95	0.83	0.88	556
2	0.98	0.96	0.97	1448
3	0.84	0.83	0.83	162
4	1.00	0.99	1.00	1608
accuracy			0.99	21892
macro avg	0.95	0.92	0.94	21892
weighted avg	0.99	0.99	0.99	21892

Figure 3.2 Classification Report of GRU with Attention over 26 Epoch

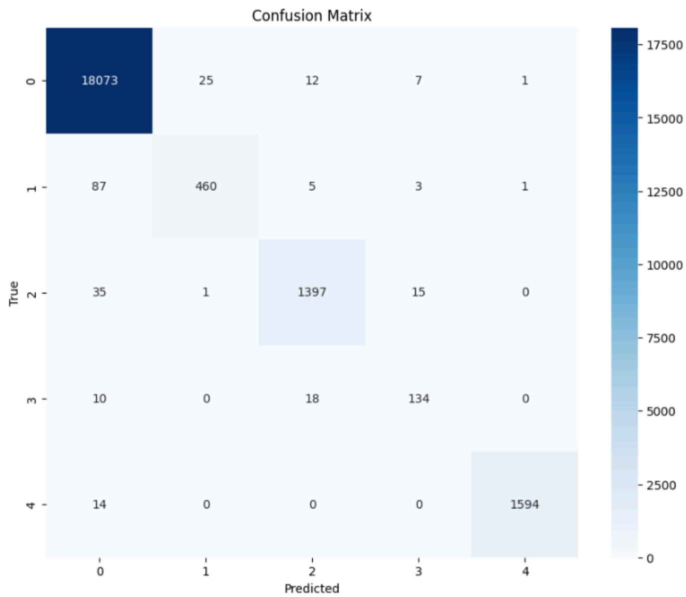


Figure 3.3 Confusion Matrix of GRU with Attention over 26 Epoch

```
Evaluating AttentionRNN model...
Testing: 100%|██████████| 685/685 [00:00<00:00, 708.29it/s]

Test Results:
Total samples: 21892
Correct predictions: 21658
Test Accuracy: 98.93%
```

Figure 3.4 Test Accuracy of CNN Model (98.93%)

Figure 4.1 Additive Noise Injection

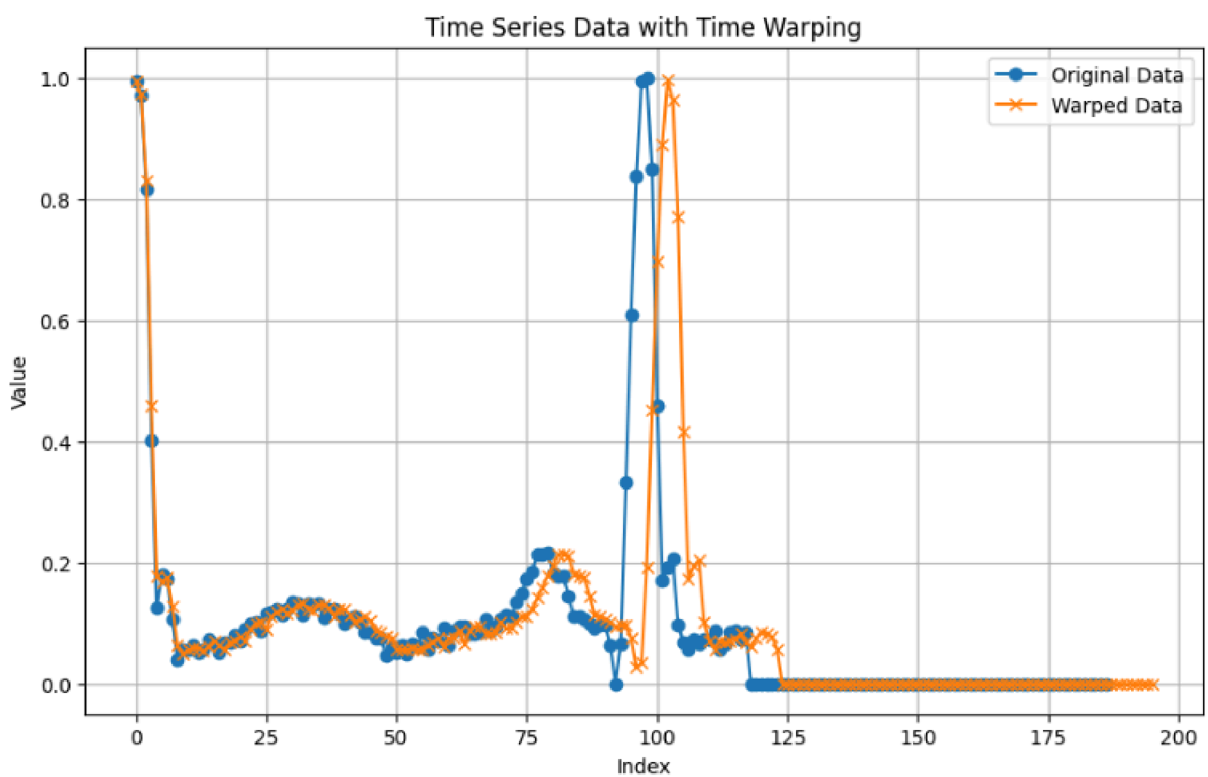


Figure 4.2 Time Warping

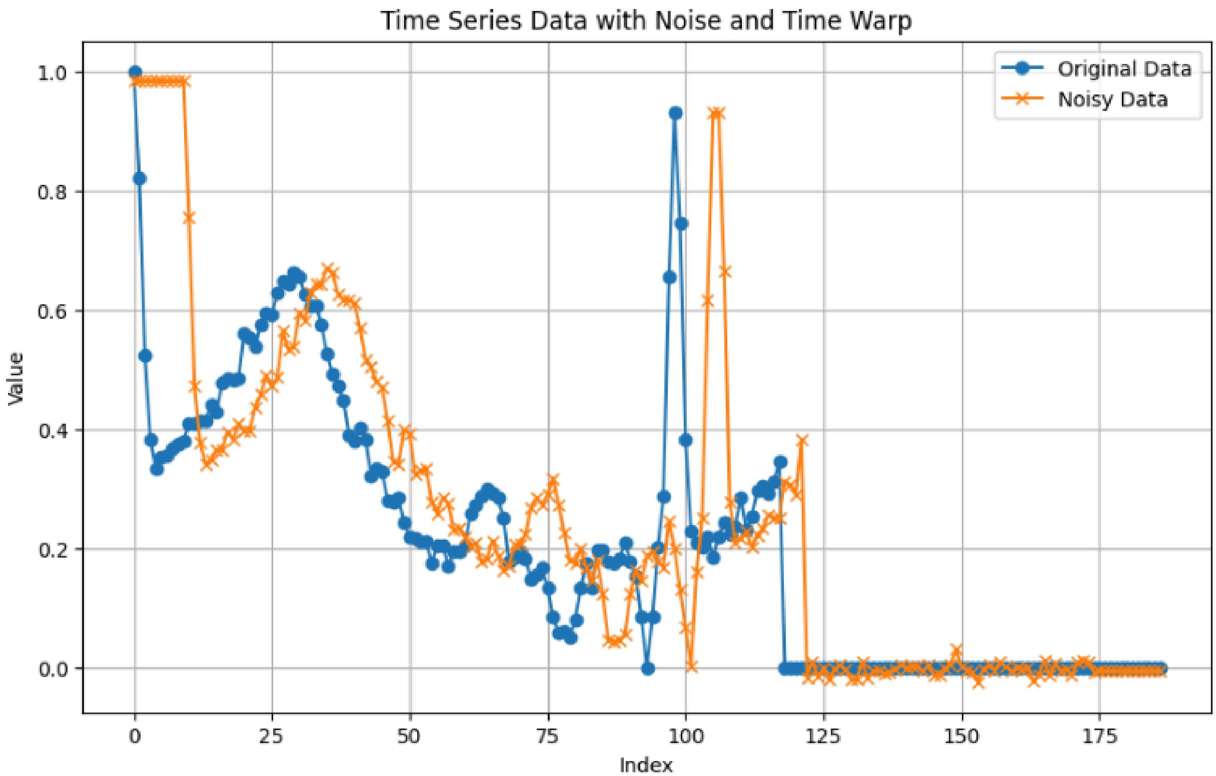


Figure 4.3 Noise Injection with Time Warping

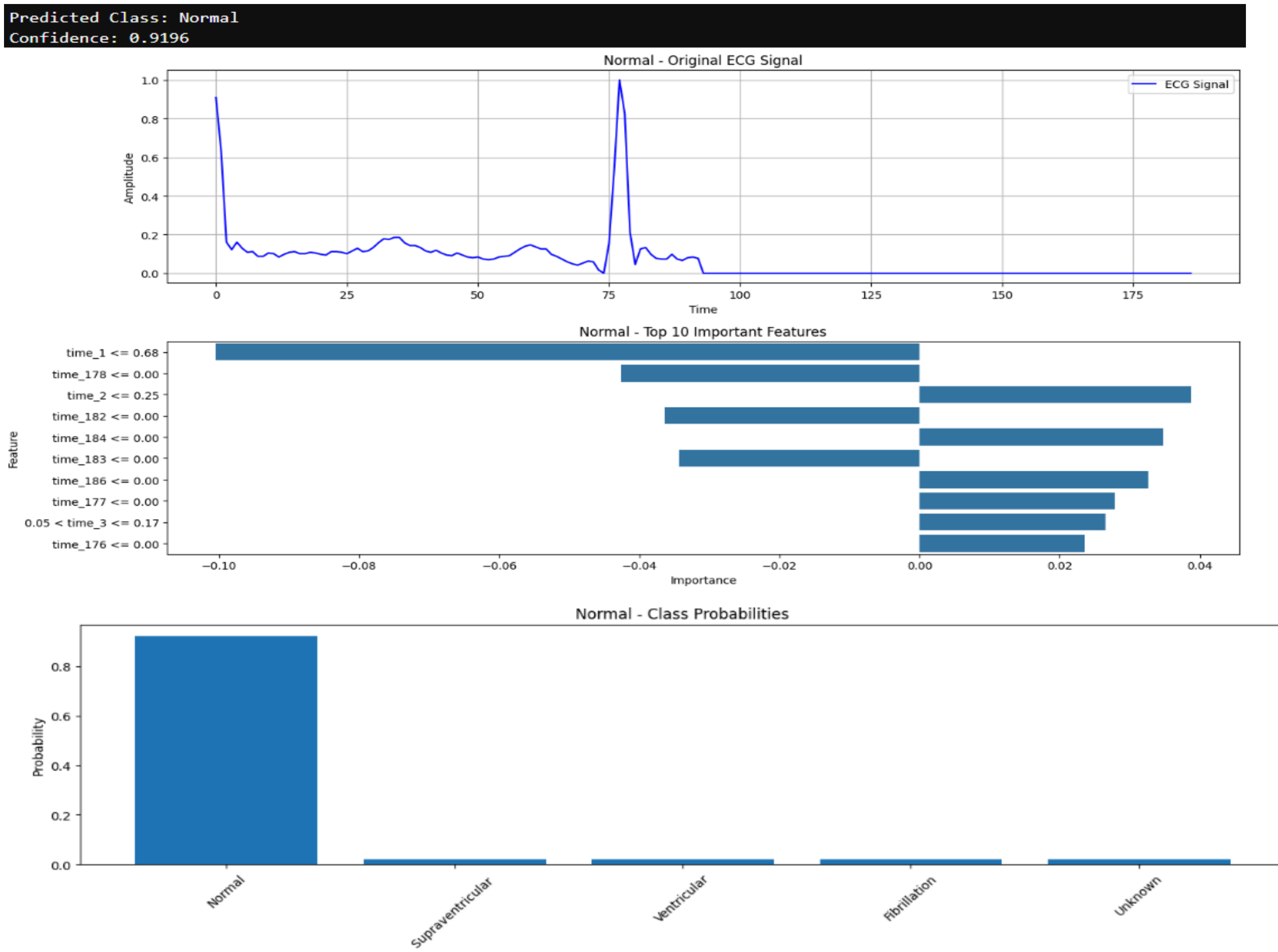


Figure 5.1 LIME Results for Class [‘Normal’] using RNN Model

1. **Top Graph** - Original ECG Signal:
    - Shows a normal ECG waveform
    - Clear QRS complex around time point 75
    - Regular baseline with normal morphology
2. **Middle Graph** - Top 10 Important Features: Most influential features with positive importance:
    - time_2 ≤ 0.52
    - time_154 ≤ 0.06
    - time_156 ≤ 0.06
    - time_153 ≤ 0.06 These features likely correspond to key points in the normal ECG morphology that help identify it as normal.
3. **Bottom Graph** - Class Probabilities:
    - Shows the model's confidence in its classification
    - Very high probability (~0.85) for "Normal" class
    - Very low probabilities for other classes (Supraventricular, Unknown, etc.)
    - The model is quite confident (91.96% as shown in the top header) that this is a Normal ECG

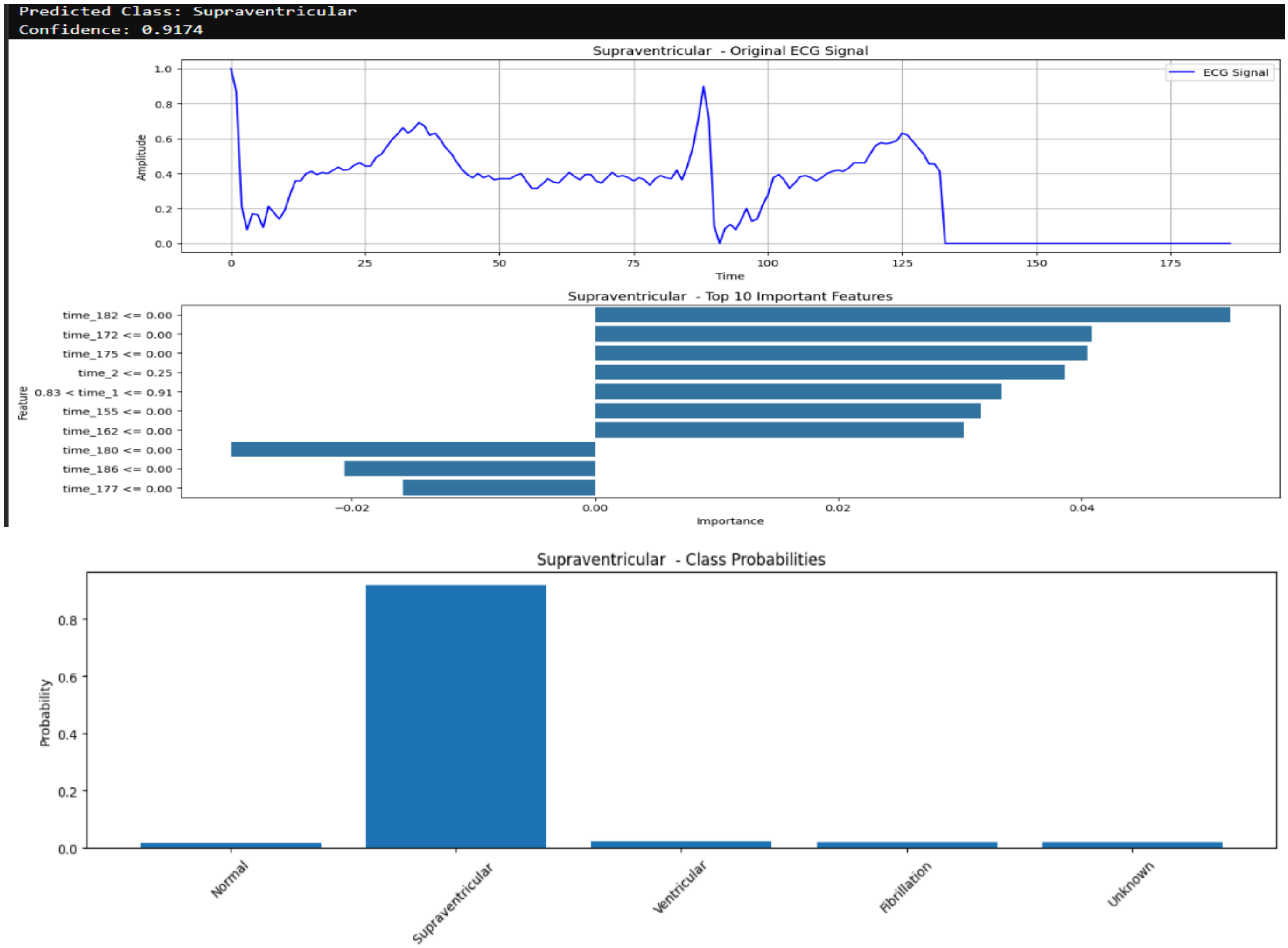


Figure 5.2 LIME Results for Class [‘Supraventricular’] using RNN Model

1. **Top Graph - Original ECG Signal**:
   - Shows a Supraventricular ECG pattern
   - Notable features include:
     - Irregular RR intervals
     - Different wave morphology compared to the normal ECG
     - Distinctive peak around time point 75 and another around 125
     - More variable baseline activity
2. **Middle Graph - Top 10 Important Features: Most influential positive features:**
   - time_183 <= 0.00 (strongest positive importance)
   - time_175 <= 0.00
   - time_2 <= 0.25
   - time_155 <= 0.00
   - time_162 <= 0.00

   Some negative features:
   - time_180 <= 0.00
   - time_177 <= 0.00
3. **Bottom Graph - Class Probabilities:**
   - Very high probability (~0.85) for "Supraventricular" class

- Low probabilities for all other classes (Normal, Unknown, etc.)
- Model confidence shown in header: 93.74%

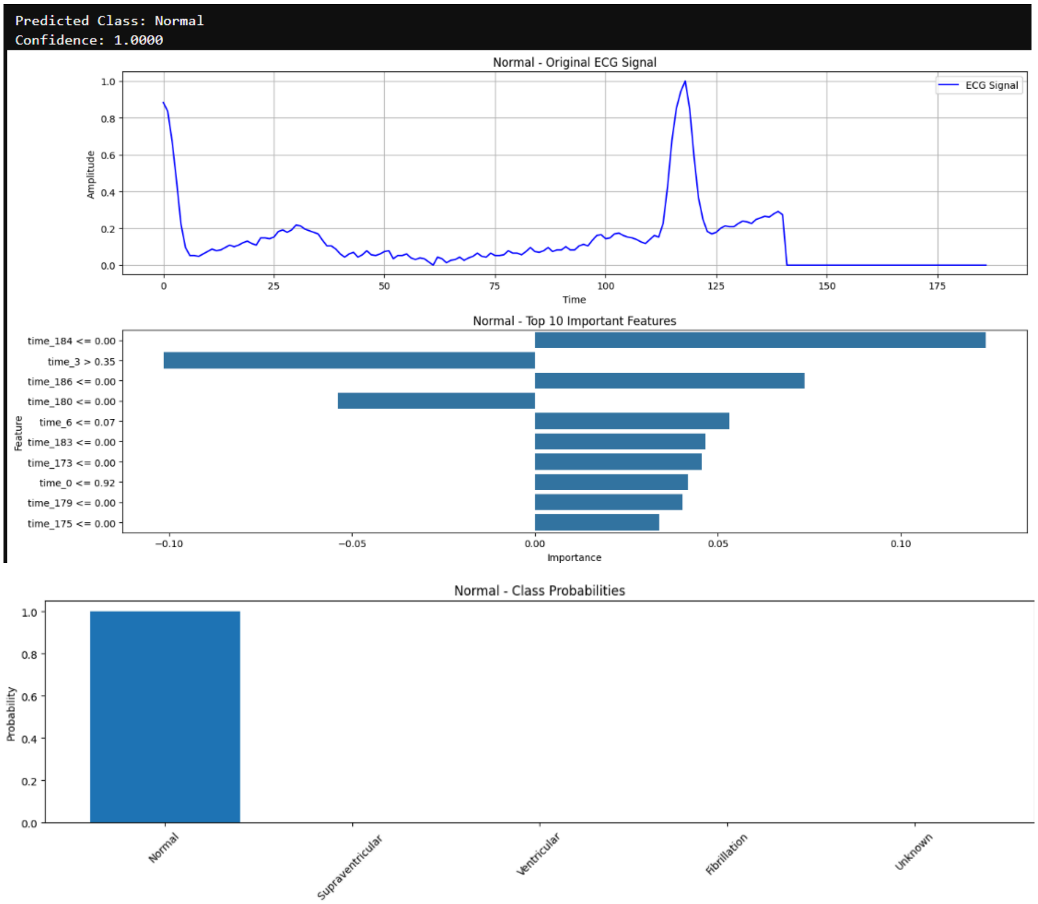


Figure 5.3 LIME Results for Class [‘Normal’] using GRU Attention Model

1. **Top Graph - Original ECG Signal:**
   - Shows a typical normal ECG pattern
   - Clear prominent R peak around time point 125
   - Stable baseline
   - Regular morphology characteristic of normal cardiac rhythm
2. **Middle Graph - Top 10 Important Features: Most significant features:**
   - time_184 <= 0.06 (highest positive importance)
   - time_108 <= 0.06 (strong positive importance)
   - time_5 ≤ 0.35
   - time_171 <= 0.06
   - time_183 <= 0.06

Some negative contributing features:

   - time_200 <= 0.06
   - time_6 <= 0.07
3. **Bottom Graph - Class Probabilities:**
   - Nearly 100% probability for "Normal" class
   - Virtually zero probability for other classes (Supraventricular, Unknown, Premature, etc.)
   - Model confidence shown in header: 1.0000 (100%)

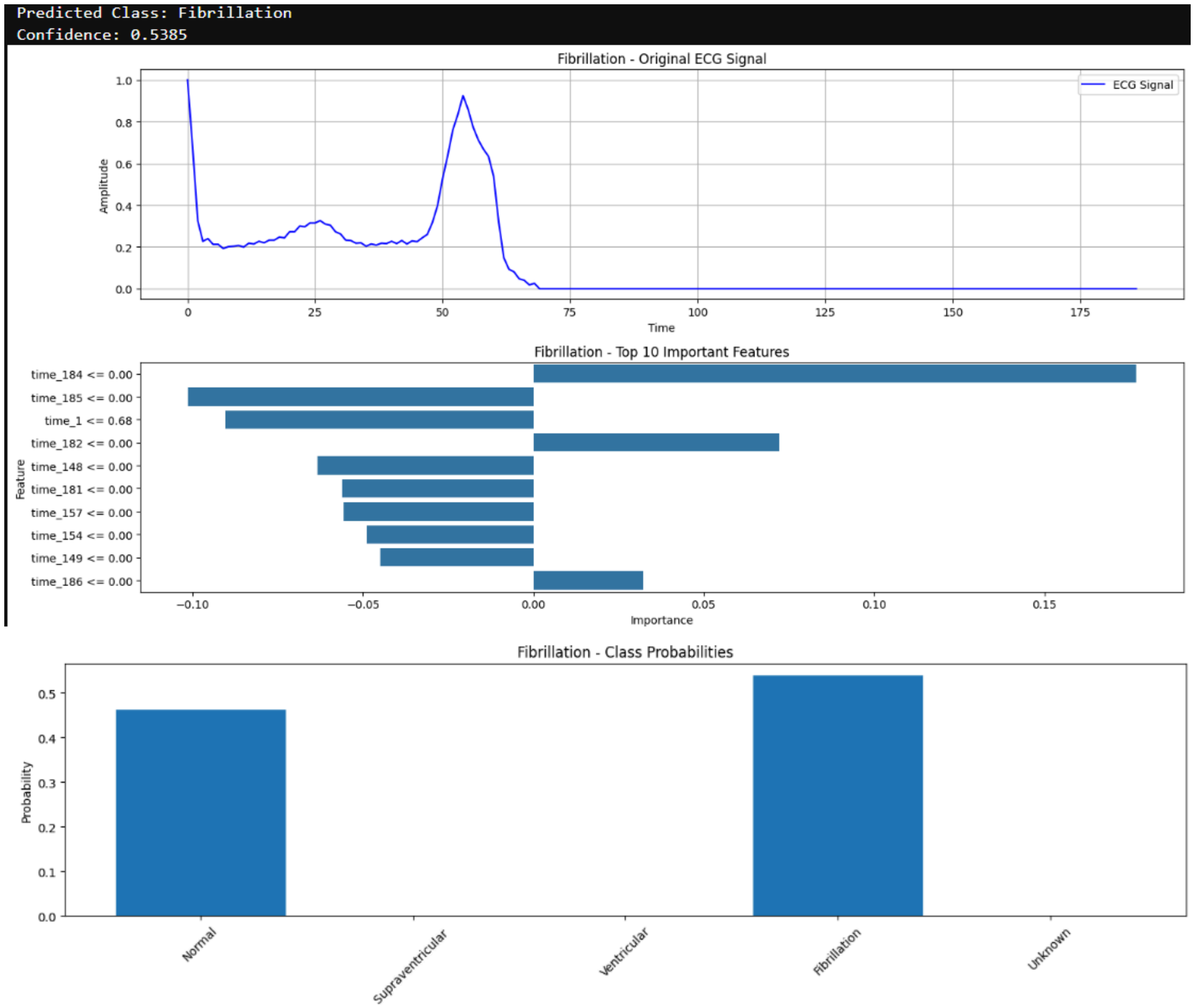


Figure 5.4 LIME Results for Class ['Fibrillation'] using GRU Attention Model

The top graph shows the original ECG signal with characteristic waves and peaks, while the bottom graph shows the LIME feature importance analysis.

The features shown are time points in the ECG signal, where:

1. **Most Important Positive Features (bars extending right):**
   - time_177 <= 0.00 (highest positive importance)
   - time_2 <= 0.55
   - time_166 <= 0.00
   - time_165 <= 0.00
2. **Most Important Negative Features (bars extending left):**
   - time_178 <= 0.00
   - time_180 <= 0.00
   - time_173 <= 0.00
   - time_183 <= 0.00

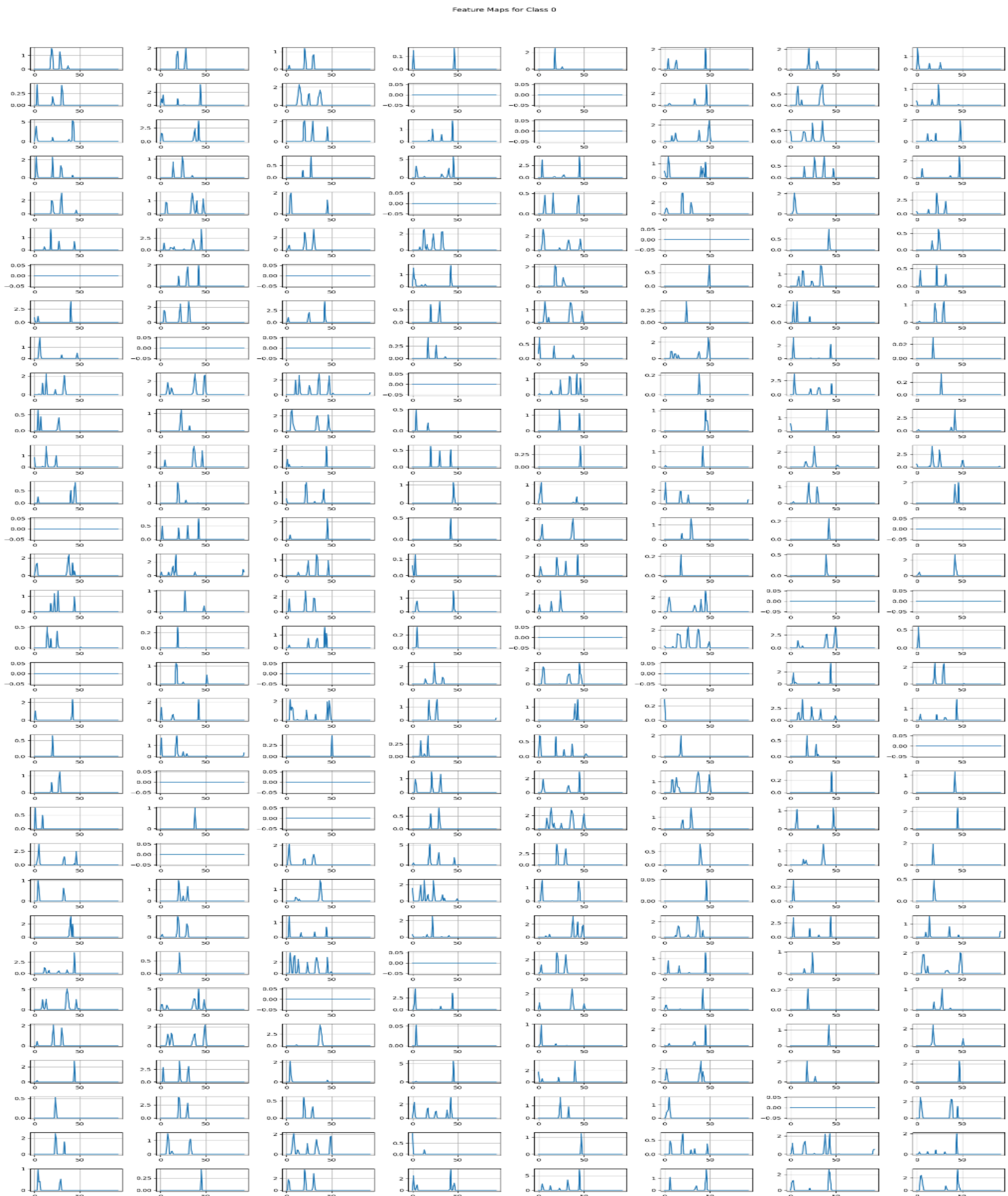


Figure 6.1, Feature Map for Class 0

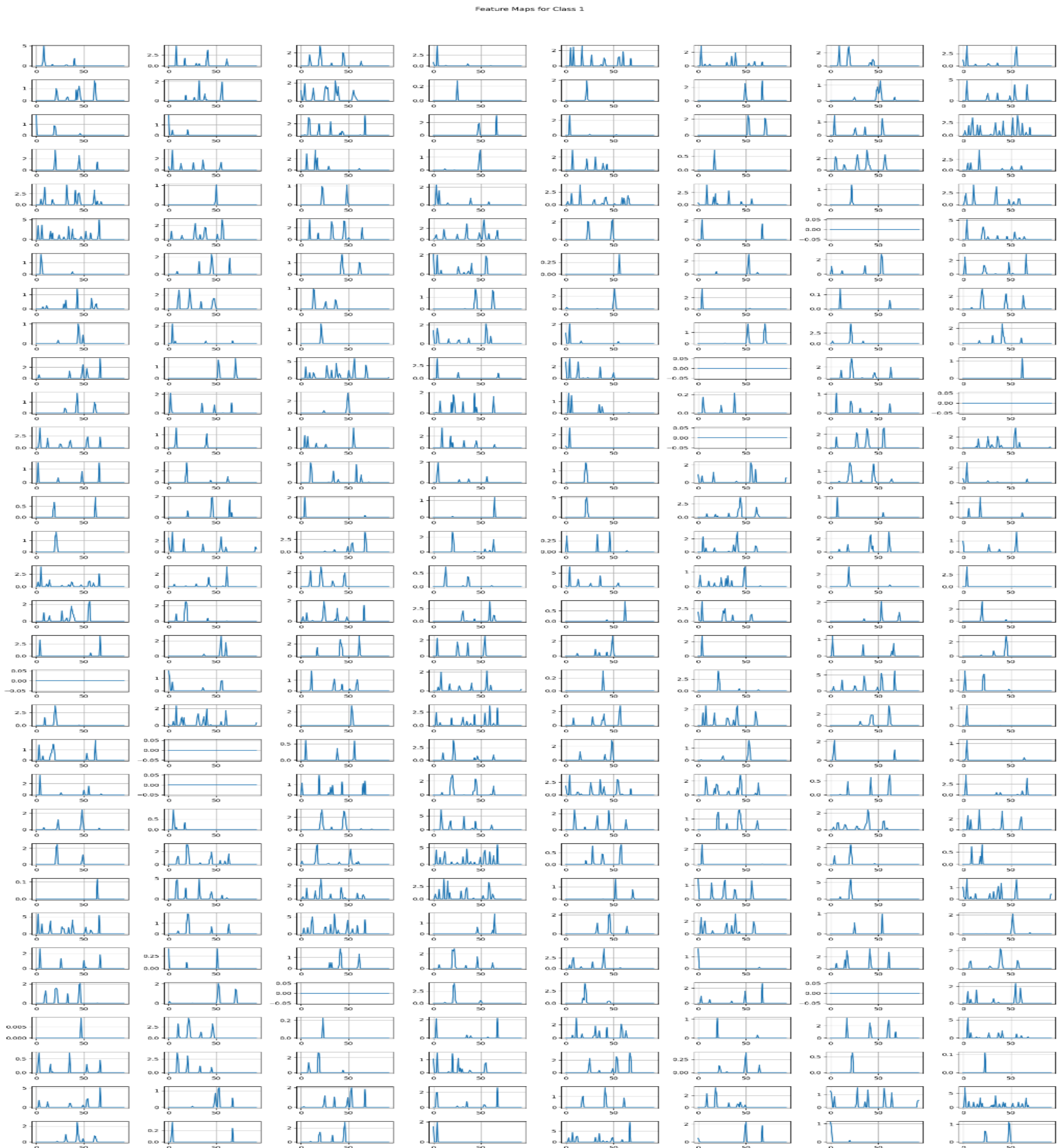


Figure 6.2, Feature Map for Class 1

Figure 6.3, Feature Map for Class 2

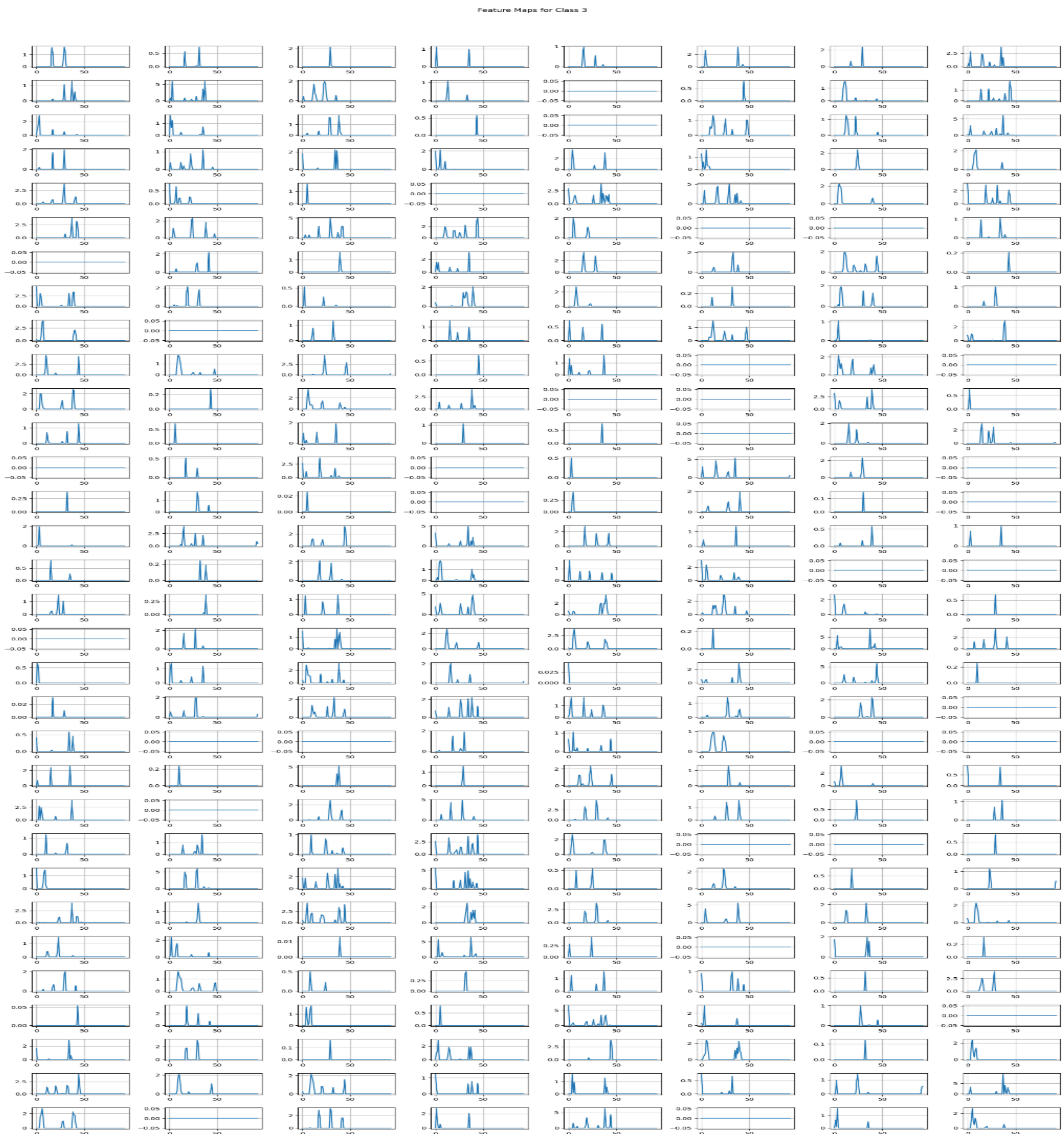


Figure 6.4, Feature Map for Class 3

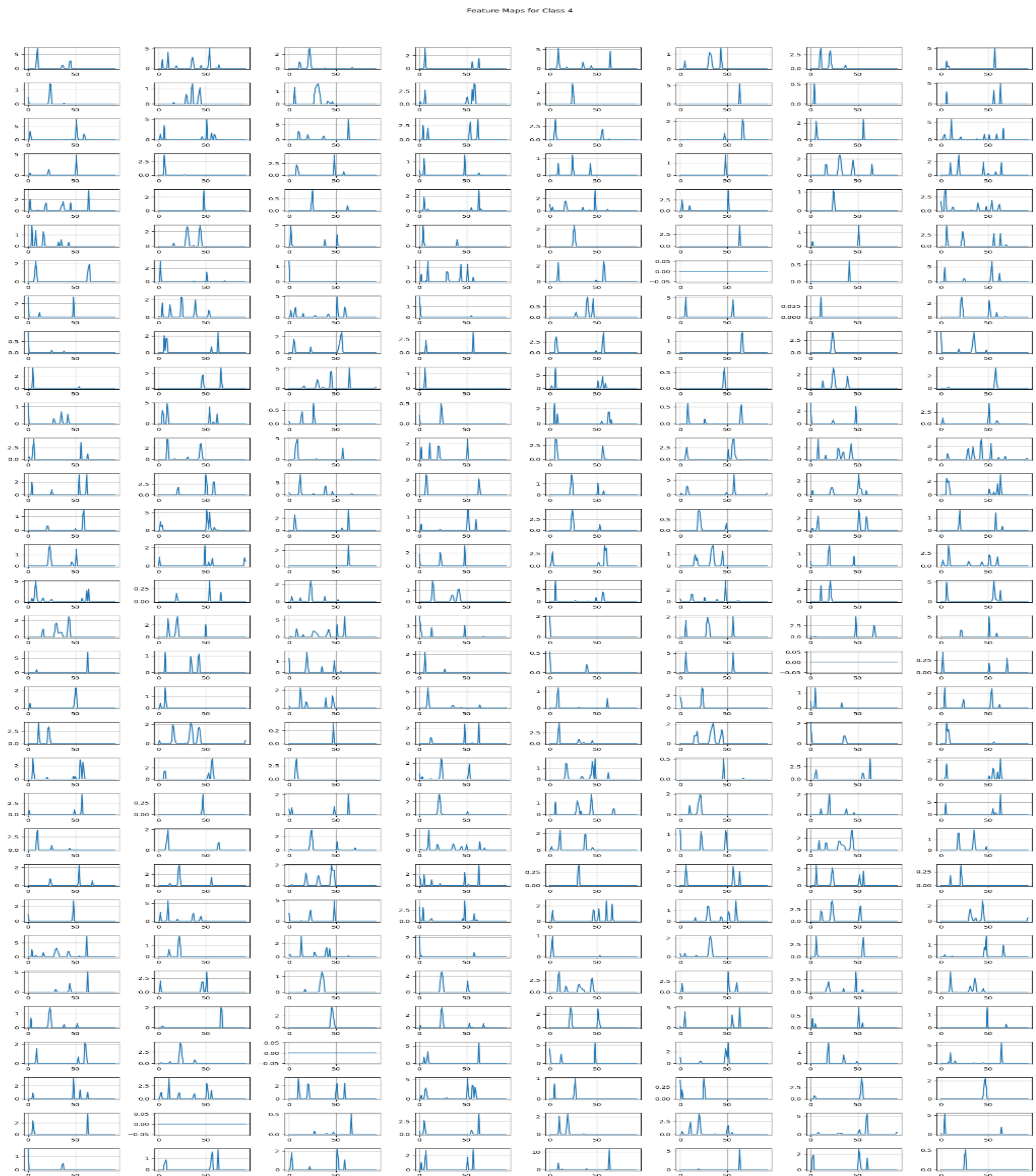


Figure 6.5, Feature Map for Class 4