**2.4 Subsystem in server**

**2.3.1 Filtering Subsystem**

**Description:**

Our model classifies the road to be safe or not according to the abnormal pattern from accelerometer. However, some abnormal pattern can be caused by the acceleration, deceleration, and turning, which is unrelated to road quality. To eliminate the influence caused by these actions, a filter is applied here. According to article written by Basavaraju [4], an 11th-order Butterworth high-pass filter with a cut-off frequency of 2 Hz can be applied to solve this problem.

**Requirement:**

1. **Exclude Unrelated Factor**

The abnormal pattern in the signal sequence should not be caused by the activity unrelated to road quality.

**Verification:**

We will get the signal from accelerometer when the vehicle does acceleration, deceleration, and turning. Compare the signal sequence before and after applying the filter.

1. **Keep Related Factor**

The abnormal pattern caused by the potential road quality problems should not be eliminated.

**Verification:**

We will get the signal from accelerometer when the vehicle run on the convex and concave road. Compare the signal sequence before and after applying the filter.

**2.3.2 Segmentation Subsystem**

**Description:**

1. **Split Into Segments:**

We need to report the road segment to have problems or not instead of the whole road, so we need to split the signal sequence into segments. To locate the problem in more precise and convenient condition, we need to make sure every signal segment is corresponding to the road segment with the same length.

1. **Preprocess:**

If all the segmentations will be processed by the Classification Module, it will create great load to system. We need to exclude the segments, which is unlikely to be related to bad road quality with a low-cost method. A threshold of speed and RMS of segment can be applied, according to [5].

**Requirement:**

1. **Keep Same Distance:**

Every signal segment should be corresponding to the road segment with the same length. An algorithm that converts the signal from GPS and accelerometer to speed information is necessary. Then the sample rate of accelerometer and speed information will be used to calculate the length of road segment.

**Verification:**

We will record the speed and distance of running vehicle. Then use the information from GPS and accelerometer to calculate speed and distance. Verify the accuracy of the algorithm by comparing the recorded value and calculated value.

1. **Correct Exclusion**

The preprocess system, which excludes the segments unlikely to be related to bad road quality, should exclude negative sample as much as possible, in the condition that no positive sample will be excluded.

**Verification:**

The threshold should make the False negative to be 0. In this situation, test different threshold that has the greatest precision.

1. **Limited Time**

The load of Segmentation Module should not be high. In other word, the preprocess algorithm should not be too complex.

**Verification:**

We will compare the time cost before and after applying the exclusion algorithm. The time cost should be significantly reduced.

**2.3.3 Classification Subsystem**

**Description:**

A machine learning based classification nodule is applied to classify the road to be safe or not. For every segment from segmentation Module, the model should classify the segment as accurately as possible. Then it will label the location information from GPS to having a problem or not.

**Requirement:**

1. **Significant Feature Extraction:**

A method of feature extraction should be chosen to make the abnormal pattern can be more easily recognized by model.

**Verification:**

We will try different methods of feature extraction to signal from the accelerometer. The abnormal pattern should be obvious from the view of people.

1. **Sufficient Training Set:**

The training set, which has high quantity and variety should be collected.

**Verification:**

We will make sure the number of samples is sufficient. Temperate requirement is 5000 segments with correct label should be collected. In addition, the signal cause from various type of road quality problems should be collected. Temperate requirement is 10 types of road quality problems should be tested.

1. **High Accuracy:**

The model should have high accuracy.

**Verification:**

We will test the accuracy of the model. The high recall is the main target. It should be over 90%. In this situation, precision should be as high as possible.

1. **Limited Time:**

The model should run in a limited time.

**Verification:**

Every segmentation should be processed within 0.1 s. In this situation, a test set with 1000 sample should be processed by model within 100s.

**2.3.4 Display Subsystem**

**Description:**

The result reported from classification subsystem is the label and the information recorded from GPS, such as longitude and latitude. These groups of information are not human-readable enough. We will need to display the information in a more readable way.

The longitude and latitude will be transformed into nodes on the map. The label from classification subsystem will be marked as the color of the nodes. In this situation, people can easily know which road segments is likely to have road quality problems.

**Requirement:**

The mapping from longitude and latitude to nodes need to be accurate.

**Verification:**

We will get the map of Haining as the map to be marked. Then we will carry our GPS sensor to run along the roads in Haining. We will choose several road segments to test. Get the signal from GPS sensor and use GPS app from phone to mark where we record the signal. Use our display subsystem to see whether the node calculated from signal of GPS sensor match the node we recorded by GPS app from phone.