**5.Experiment and Discussion**

**5.1 experiment for UCEkNN and optimal k value selection**

To testify our proposed two models(UCEkNN, optimal k value selection), we conduct two experiments which is **experiment I** (for UCEkNN) and **experiment II(**for optimal k value selection**)** using the same real data set in the MdtDSS mentioned in [9]. The size of the dataset is 3340 if we only used it to examine the general kNN Algorithm, however, only1455 of them contain the voting information which can be used in our uncertainty based method.

In the 1455 instances which have OpS, the average size of OpS is 8 which means there are 8 doctor voting to one patient on average. We repeat the two experiment 3 times respectively based on 3 different subsets of the 1455 instances. Each of the subsets are divided into 3 segments which have different function. The 3 segments are training data set, verification data set and testing data set. Experiment I and II use different part of the 3 segments: (1)experiment I use the training data set (2) experiment II using the training data set, verification data set and testing data set at the same time. The length of each segment during the 3 reduplicative experiments is illustrated as follows:

| Experiment number | training | verification | testing |
| --- | --- | --- | --- |
| 1 | 1028 | 212 | 215 |
| 2 | 790 | 238 | 427 |
| 3 | 610 | 347 | 239 |

In **experiment I**, we predict the label of each instance in the testing data set using two method (1) the usual DS evidence rule based kNN Algorithm, and (2) our proposed uncertainty based kNN algorithm (UCEkNN). In this experiment, we use the globally optimal k as the input of these two algorithms. First of all, we devise a simple greedy algorithm: select k in [kmin,kmax] successively and calculate the predicting accuracy with our proposed UCEkNN, and the optimal k is who has the best accuracy. Whereafter, we compare the forecast result and the actual result to judge if the prediction is correct, at last we calculate the accuracy of this experiment. The results of all these 3 reduplicative experiments are as follows:

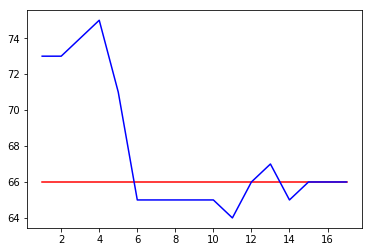
| Experiment number | EkNN | UCEkNN |
| --- | --- | --- |
| 1 | 59.53% | 66.05% |
| 2 | 63.00% | 66.04% |
| 3 | 58.16% | 64.01% |

In **experiment II**, we want to exhibit the improvement we made compared to a fixed globally optimal k if we use different k acquired by our proposed method. Thus, we use (1) the fixed globally optimal k (2) different k acquired by the **Algorithm 2** (3) different k acquired by the **Algorithm 3** separately to predict the label of each instance in testing data set based on UCEkNN and compute the accuracy, we get:

| Experiment number | fixed k | diff k based on Alg.1 | diff k based on Alg.2 |
| --- | --- | --- | --- |
| 1 | 66.05% | 70.70% | 69.30% |
| 2 | 66.04% | 69.79% | 69.09% |
| 3 | 64.01% | 67.36% | 67.78% |

The **experiment II** felicitously simulate a realistic scene where the kNN is used to predict a new pattern. In such a scene a unlabeled instance get its label then become a historical instance and turn into a part of the training set. It is inflexible and unadvisable to keep a fixed k nor calculate a new optimal k at a regular interval. In contrast, our proposed method get a good performance in the accuracy and can saving huge computing resources.

**5.2 experiment for optimization based on L-Sure algorithm**

To verify the efficiency of our proposed L-Sure algorithm and intensify the comparison against the normal UCEkNN algorithm, we conduct an experiment using the same data set with experiment 2, and the capacity of the training, verification, and testing data set is 610,347,239 respectively. In our experiment, the value of L varies in the range of 0 to 16, and the result is as follows:

the red curve represents the normal UCEkNN while the blue curve represents UCEkNN with L-Sure algorithm. This figure shows that when L meets 6 to 11, the error times is lower than that without using L-Sure algorithm.