



## Multisample Classification in Clinical Decisions using Multi-Aggregative Factored K-NN Classifier

**P.Tamil Selvan**

Research Scholar, Hindusthan College Of Arts &  
Science, Coimbatore, Tamil Nadu

**Dr. Senthil Kumar A.V**

Research Supervisor, Hindusthan College Of Arts &  
Science, Coimbatore, Tamil Nadu

### ABSTRACT

Classification in sample by sample process, a classifier is requested to combine information across multiple samples drawn from the same data source, the results are combined using a strategy such as majority are selected. To solve the problem of classification failure, a new hazard function in multisample classification is introduced ie Multi-aggregative factored K-NN Classifier. This method evaluates the classification of multisampling problems, such as electromyographic (EMG) data, by making aggregate features available to a per-sample classifier. It is found that the accuracy of this approach is superior to that of traditional methods such as majority selection for this problem. The classification improvements of this method, in conjunction with a confidence measure expressing the per-sample probability of classification failure (i.e., a hazard function) is described and measured. This paper compares the existing method Bayesian and the proposed Multi-aggregative factored KNN approach. The experimental results displayed a prominent improvement by using the proposed algorithm.

**Keywords:** EMG, Motor Unit Action Potential, Additional Feature Sets, Classifier

### 1. INTRODUCTION

Electromyographic Signal is introduced as an element of time and can be portrayed regarding its amplitude, frequency and phase. It is a signal which measures electrical streams produced amid the withdrawal of the muscle which speaks to the neuromuscular action of that muscle. Out of three sorts of muscles in human

body, EMG signals are gathered from skeletal muscles [1]. The skeletal muscle tissues are joined to the bones. The withdrawal of these muscles is in charge of supporting and moving the human skeleton. At the point when a drive is produced by the neuron, the withdrawal of the skeletal muscle is started which is typically intentional [2]. Skeletal muscles are examined so as to acquire the EMG information. Skeletal muscle filaments have bounteous neurons for its compression. This kind of neuron which produces compression are called engine neurons and are put nearer to muscle tissue however not really associated with it. One engine neuron can give incitement to many muscle strands. The human body is considered electrically unbiased all in all as it has a similar number of positive and negative charges. However, while in the resting state, the nerve cell layer is enraptured because of contrasts in the fixations and ionic creation over the plasma film.

### 2. EXISTING SYSTEM

Several Bayesian classifiers were compared. PD/FIS\*, a rule-based classier, and three Bayesian networks: naive Bayes (NAIVE-BN), tree-augmented naive Bayes (TAN-BN) and an evolutionarily constructed Bayesian network (EVOLVED-BN). PD/FIS-This classifier has already been utilized with QEMG information. It works by assessing the recurrence of events of relationship between estimations of the mark and watched includes in at least one of the information sections [3]. By looking at these, utilizing the balanced lingering one may recognize affiliations that contrast fundamentally from those normal by a

model of irregular possibility; these "examples" are then utilized as standards for classification, weighted by their data content utilizing the "event/all" component.

A Bayesian network (BN) is a coordinated non-cyclic chart based portrayal of a likelihood dissemination, utilizing hubs to speak to perceptible occasions, for example, specific info esteems or class names, and relations between occasions as bends. Hunting down an ideal chart in light of preparing information is troublesome, both because of the need to set up the level of reliance between watched occasions, and the computational many-sided quality of the inquiry [4]. We look at three common calculations for acquiring a non optimal chart in a plausible way. Certainty might be measured for every single Bayesian network by looking at the part of probabilistic help for the triumphant class: this division is then utilized as the trust in the relegated classification, for C NAIVE-BN, CTan-BN, and CEvolved-BN. Naive Bayesian networks (NAIVE-BN) in view of the suspicion of finish autonomy between input esteems, are shockingly powerful classifiers, often beating more mind boggling classifiers. A vital shortcoming of NAIVE-BN in CDSS configuration is that an inability to precisely mirror the likelihood appropriation of the fundamental information prompts a poor measure of choice certainty, and undermines straightforwardness and understand ability [5-6].

Tree augmented Naive Bayesian Networks (TAN-BN) to misuse the qualities of NAIVE-BN classifiers by unwinding the autonomy presumption, permitting the element hubs in a network to frame a completely subordinate tree, making frameworks that can outflank NAIVE-BN frameworks [7]. Evolutionary Algorithms Additionally using randomized hunt, a transformative calculation can be utilized to develop the network, by utilizing competition choice randomization to choose networks for consolidating, and by swapping circular segments, lastly pruning by the utilization of a Markov cover in view of the class hub as portrayed. Using Bayesian learning systems, we evaluate the efficacy of using additional feature sets (AFSs) on MUP data, where an overall muscular characterization is required based on the "study" of the problem, with multiple samples drawn from the same source[8-11]. Some further exploration of these ideas using studies drawn from synthetically generated covaried data were also performed. Values for an individual AFS are calculated by using a simple

aggregation of all of the observed values for each feature within the study, and adding this result as a new feature to all samples, providing each sample information about the entire study. We inspect three simple aggregators in this initial examination of this idea: arithmetic mean, and maximum and minimum value [12-15].

Bayesian Learning is relevant in explicit manipulation of probabilities among the most practical approaches to certain types of learning problems. e.g. Bayes classifier is competitive with decision tree and neural network learning.

```
model = GaussianNB()
# Train the model using the training sets
model.fit(x, y)
#Predict Output
predicted= model.predict([[1,2],[3,4]])
print predicted
require(e1071) #Holds the Naive Bayes Classifier
Train <- read.csv(file.choose())
Test <- read.csv(file.choose())
levels(Train$item_Fat_Content)
model <- naiveBayes(Item_Fat_Content~, data =
Train)
class(model)
pred <- predict(model,Test)
table(pred)
```

### Algorithmic Process for Existing Methods

The second reason: useful perspective for understanding learning methods that do not explicitly manipulate probabilities determine conditions under which algorithms output the most probable hypothesis e.g. justification of the error functions in ANNs e.g. justification of the inductive bias of decision trees [16-18]. Each observed training example can incrementally decrease or increase the estimated probability that a hypothesis is correct. Prior knowledge can be combined with observed data to determine the final probability of a hypothesis. Hypotheses make probabilistic predictions new instances can be classified by combining the predictions of multiple hypotheses, weighted by their probabilities. Standard of optimal decision making against which other practical measures can be measured.

### 3. PROPOSED SYSTEM

The proposed system is Multi-aggregative factored K-NN Classifier for classification of multiple samples. Five aggregative factors are considered for aggregating features, i.e. Best Feature value, Worst Feature Value, arithmetic mean, and maximum and minimum value. All information were quantized utilizing greatest peripheral entropy all together that Bayesian occasion probabilities might be developed on the information as quantized into ten canisters. Forget one cross approval was utilized to better gauge classification precision, utilizing each total investigation as a solitary cross-approval set; this will guarantee that the greater part of the related AFS esteems from each examination are assembled together into either testing or preparing datasets.

In this paper we proposed an enhanced Multi-aggregative factored KNN approach, which is denoted as KNN++, for classifying complex data with heterogeneous views. Any type of view can be utilized when applying the KNN++ method, as long as a distance function can be defined on that view. The KNN++ includes an integral learning component that learns the weight of each view. Furthermore, the KNN++ method factors in not only the training data, but also the unknown instance itself when assessing the importance of different views in classifying the unknown instance. An Enhanced KNN Approach for Classification Given a set of data instances  $U$  with  $N$  elements  $\{u_1; u_2; \dots; u_N\}$  and a set of class labels  $C$  with  $M$  elements  $\{c_1; c_2; \dots; c_M\}$ ,  $U$  is divided into  $M$  disjoint regions  $\{rc_1; rc_2; \dots; rc_M; rc_M \neq 1\}$ , such that if a data instance  $u_i \in rc_j$  (where  $1 \leq j \leq M$ ), then the class label  $c_j$  is assigned to  $x_i$ ; if  $u_i \notin rc_M \neq 1$ ,  $u_i$  is viewed as an unknown instance. Now, the classification problem that is addressed here is that, for each  $u_i \in rc_M \neq 1$ , we need to assign a class label  $c_j \in C$  to it. We further assume that a set of distinct distance functions  $D = \{d_1; d_2; \dots; d_L\}$  can be defined on  $U$ , such that for any  $d_x \in D$  and any  $u_i; u_j; u_k \in U$ ; we have  $d_x(u_i; u_j) \leq d_x(u_j; u_k) \leq d_x(u_i; u_k)$ . The classical KNN approach assumes that  $j \in D$ ; in other words, only one distance function is used in the classification process. However, a complex data set may have multiple heterogeneous views. It is often challenging, if not possible, to define one single comprehensive distance function that is able to take into consideration of multiple heterogeneous views. Therefore, the proposed KNN++ method utilizes multiple distance functions, each of which is defined on one heterogeneous view of the data. Let's take as

an example the data-driven detection of Alzheimer's disease based on patient data. One distance function on patient cases may be defined on brain images; one distance function may be defined on patients' genetic risk profiles; another distance function may be defined on trajectories of certain biomarker; and so on. In this case, it is obviously difficult to define one single distance function based on all of these heterogeneous views. However, different distance functions may be defined on different views of the given data, such that one distance function represents the view upon which the function is defined. Therefore, in order to take advantages of multiple views, the proposed KNN++ method utilizes multiple distance functions. We also need to consider that not every view of the data has equal significance towards the classification of a given instance. Therefore, an important component of this proposed KNN++ method is to learn the weight of each distance function that is defined on each view. Furthermore, the weights of distance functions should not remain unchanged for different unknown instances. For instance, given certain patient case, brain image may be more important than others in detecting the disease; while for another case, a biomarker may serve as a better indicator. Hence, the learning process of the proposed KNN++ method is instance based. In other words, different unknown instances may favor different views. Informally, the KNN++ method can be described in the following way. Given an unknown instance, the method first learns the weight of each distance function that KNN++: An Enhanced K-Nearest Neighbor Approach is defined on each view of the data.

*Let  $(X_i, C_i)$  where  $i = 1, 2, \dots, n$  be data points.  $X_i$  denotes feature values &  $C_i$  denotes labels for  $X_i$  for each  $i$ . Assuming the number of classes as 'c'  $C_i \in \{1, 2, 3, \dots, c\}$  for all values of  $i$*

*Let  $x$  be a point for which label is not known, and we would like to find the label class using  $k$ -nearest neighbor algorithms.*

1. Calculate " $d(x, x_i)$ "  $i = 1, 2, \dots, n$ ; where  $d$  denotes the Euclidean distance between the points.
2. Arrange the calculated  $n$  Euclidean distances in non-decreasing order.
3. Let  $k$  be a +ve integer, take the first  $k$  distances from this sorted list.
4. Find those  $k$ -points corresponding to these  $k$ -distances.
5. Let  $k_i$  denotes the number of points belonging to the  $i^{\text{th}}$  class among  $k$  points i.e.  $k \geq 0$
6. If  $k_i > k_j \forall i \neq j$  then put  $x$  in class  $i$ .

lights and so on. The surfaces of our bodies are continually presented to electromagnetic radiation and it is difficult to stay away from this introduction. The overwhelming worry for the surrounding commotion emerges from the 50 Hz (or 60 Hz) radiation from control sources. Motion artifacts have two fundamental wellsprings of movement curio: one from the interfacing layers between the location surface of the cathode and the skin; the other from development of the link associating the terminal to the enhancer. The electrical signs of both clamor sources have the greater part of their vitality in the recurrence extend from 0 to 20 Hz.

### Multi-aggregative factored KNN Algorithm

The weight of a distance function is determined by the labelled representatives of the unknown instance with respect to this distance function. More specifically, the  $K$  nearest neighbours of the unknown instance, which is found using this distance function, serves as the labelled representatives of the unknown instance corresponding to this distance function.

For each of the labelled representatives, the KNN++ method finds the  $K$  nearest neighbors of this labelled representative by using the same distance function; then counts how many instances within the  $K$  nearest neighbours of this labelled representative actually have the same class label as this representative. The weight of this distance function is then determined by summing up all those numbers across all the labelled representatives. After the weights of all those distance functions are calculated, the set of the  $K$  nearest neighbors found by each of the distance functions for the unknown instance is weighted by the weight of that distance function. That means, the class label of each instance in those sets of  $K$  nearest neighbors is weighted by the weight of the set that this instance belongs to. Then, the final class label that is assigned to the unknown instance by this KNN++ method is the one with the highest weighted sum across all the sets of  $K$  nearest neighbors of this unknown instance.

Inherent noise is the electronic segments utilized as a part of the recognition and recording of EMG signals creates electrical clamor. This clamor has recurrence segments that range from 0 Hz to a few thousand Hz which can't be wiped out. It must be decreased by utilizing amazing electronic segments, astute circuit outline and development methods. Ambient noise is commotion begins from wellsprings of electromagnetic radiation, for example, radio and TV transmission, electrical-control wires, fluorescent

### 4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed approach has been compared with the novel KNN++ Multi-aggregative Algorithm, the key metrics such as Best Feature value, Worst feature value, Arithmetic Mean, Maximum Value .the proposed approach has been implemented with j2ee platform for better simulation results. This is done by calculating the " $d(x, x_i)$ "  $i = 1, 2, \dots, n$ , Euclidean distances is calculated in non-decreasing order, finally finding the  $K$ -points to the  $K$ -Distance. The below figure demonstrates the results achieved on the simulation environment.

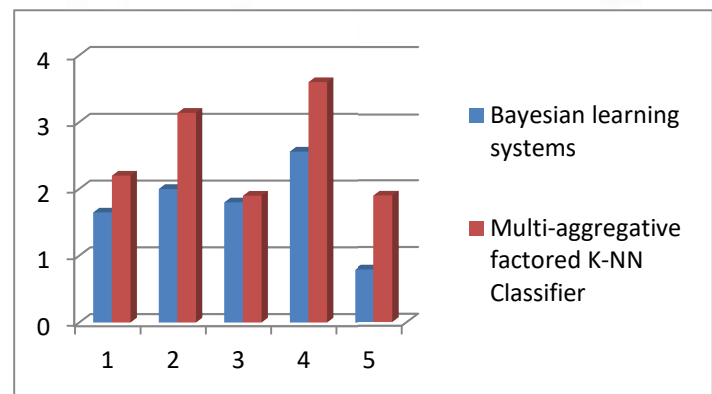
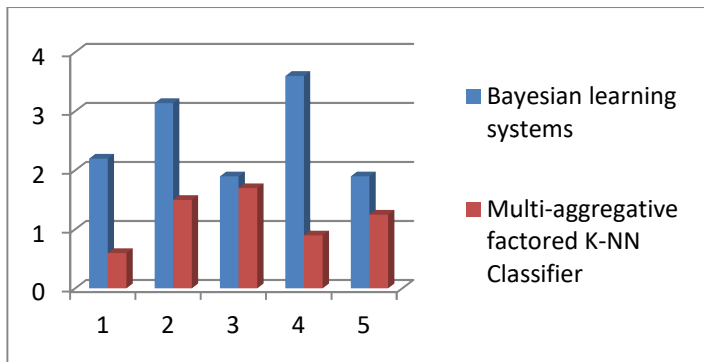


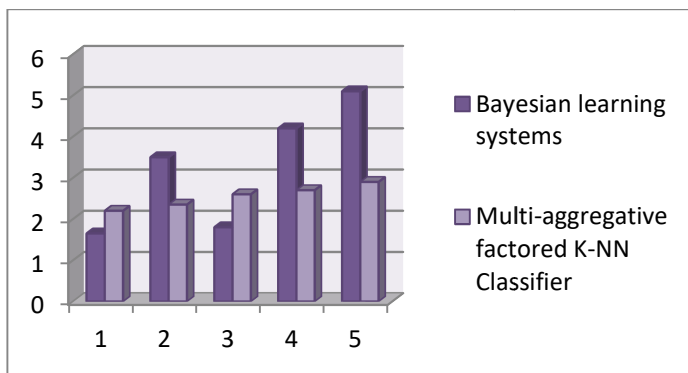
Figure 1- Best Feature value

Figure 1 display the best feature value has been gradually improved in Multi-aggregative factored KNN when compared with the existing Bayesian Learning methods. The results produced have been improved up to maximum 12% improvement and a minimum of 0.5 % improvement.



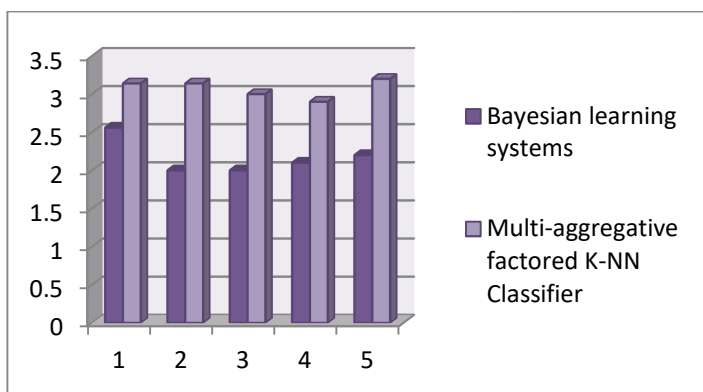
**Figure 2 - Worst feature value**

Figure 2 display the worst feature value has been decreased modularly in Multi-aggregative factored KNN when compared with the existing Bayesian Learning methods. the results produced has been improved up to an average of 0.75 % improvement .



**Figure 3- Arithmetic Mean**

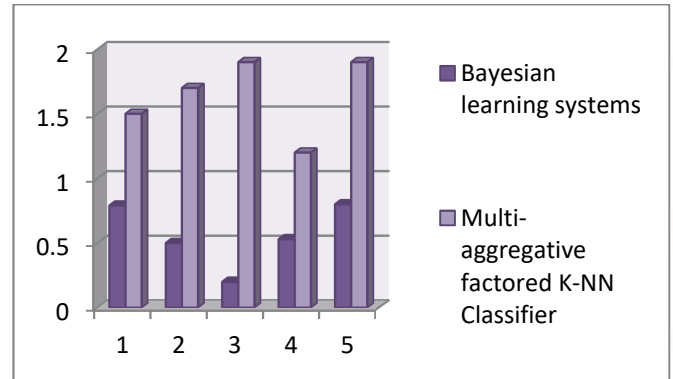
Figure 3 displays the Arithmetic feature value has been gradually improved in Multi-aggregative factored KNN when compared with the existing Bayesian Learning methods. The result produced has been improved up to maximum 15% improvement and a minimum of 1 % improvement.



**Figure 4 -Maximum Value**

Figure 4 displays the maximum value has been gradually improved in Multi-aggregative factored KNN when compared with the existing Bayesian

Learning methods. The result produced has been improved up to maximum 6% improvement and a minimum of 1.5 % improvement.



**Figure -5- Minimum Value**

Figure 5 displays the Minimum value has been gradually improved in Multi-aggregative factored KNN when compared with the existing Bayesian Learning methods. The results produced have been improved up to maximum 80% improvement and a minimum of 3 % improvement.

The results from figure 1 to 5 display a gradual improvement in Best Feature value, Worst Feature Value, arithmetic mean, and maximum and minimum value. All information were quantized utilizing greatest peripheral entropy all together that Bayesian occasion probabilities might be developed on the information as quantized into ten canisters.

## CONCLUSION

EMG signals are important in different biomedical and neurological applications. EMG signals are non-stationary and also non-uniform. They are not repeatable and every so often can even be clashing. From now on, the get ready of such signals transforms into a troublesome work. The development to record and separate the EMG signal is tolerably new. Therefore, there are various confinements in area and depiction of existing nonlinearities in the surface electromyography signal, estimation of the stage and acquiring clear information in view of assurance from normality. The expansion of AFSs to the first MUP information builds free example characterization exactness, however that does not convert into expanded examination grouping precision. Whenever prepared and tried with AFS (MEAN), the PD/FIS\* frameworks appear a decreased certainty blunder. Whenever AFS (MAX) is utilized to prepare and test the PD/FIS\* frameworks, certainty blunder stays

unaltered. This paper compares the existing method Bayesian and the proposed Multi-aggregative factored KNN approach. The experimental results displayed a prominent improvement by using the proposed algorithm

## REFERENCES

- 1) Robert Varga, S. Marie Matheson, and Andrew Hamilton-Wright, "Aggregate Features in Multisample Classification Problems "IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 2, MARCH 2015.
- 2) Raez, M.B.I., M.S. Hussain, and F. Mohd-Yasin. "Techniques of EMG Signal Analysis: Detection, Processing, Classification and Applications." *Biological Procedures Online* 8 (2006): 11–35. PMC. Web. 13 Feb. 2017.
- 3) A. Averbuch, V. A. Zheludev, N. Rabin and A. Schclar, "Wavelet-based acoustic detection of moving vehicles", *Multidimensional Systems and Signal Processing*, vol. 20, no. 1, (2009), pp. 1-25.
- 4) Harish Chander Mahendru , "Quick review of human speech production mechanisml" , *IJERD* , Volume 9, Issue 10, January 2014
- 5) Victoria-M. Florescu<sup>1</sup>, Lise Crevier-Buchman et al. " Silent vs Vocalized Articulation for a Portable Ultrasound-Based Silent Speech Interface " , *INTERSPEECH 2010* , 26-30 September 2010, Makuhari, Chiba, Japan
- 6) S.-C. Jou, T. Schultz, M. Walliczek, F. Kraft, and A. Waibel, "Towards continuous speech recognition using surface electromyography," in *Proc. Interspeech*, 2006, pp. 573–576
- 7) Matthias Janke, Michael Wand, and Tanja Schultz , "A Spectral Mapping Method for EMG-based Recognition of Silent Speech " , *Cognitive Systems Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany*
- 8) Michael Wand, Christopher Schulte, Matthias Janke, Tanja Schultz, "Array based electromyographic silent speech interface", *Cognitive Systems Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany*
- 9) Shubhangi S. Jarande , Prof. Surendra Waghmare, "A Survey on Different Classifier in Speech Recognition Techniques", *IJETAE* , Volume 5, Issue 3, March 2014
- 10) Shanthi Therese S. ,Chelpa Lingam, "Review of Feature Extraction Techniques in Automatic Speech Recognition", *IJSET* , Volume No.2, Issue No.6, pp : 479-484, June 2013
- 11) Janne Pylkkonen, " LDA Based Feature Estimation Methods for LVCSR" , *Cognitive Systems Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany* Adaptive Informatics Research Centre Helsinki University of Technology, Finland
- 12) Michael Wand, Matthias Janke, and Tanja Schultz, " Tackling speaking mode varieties in EMG based speech recognition " in *IEEE transactions on biomedical engineering*, vol. 61, no. 10, may 2014
- 13) Rupali V. Pawar, Dr. R.M. Jalnekar, " Review on speech production model", *IJEIT*, Volume 3, Issue 9, March 2014
- 14) Kirandeep Kaur, Neelu Jain, " Feature Extraction and Classification for Automatic Speaker Recognition System " A Reviewl, *IJARCS*, Volume 5, Issue 1, January 2015
- 15) N. Sugie and K. Tsunoda, " A speech prosthesis employing a speech synthesizer " Vowel discrimination from perioral muscle activities and vowel production, " *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 7, pp. 485–490, Jul. 1985
- 16) S.-C. Jou, T. Schultz, and A. Waibel, "Continuous electro myographic speech recognition with a multi-stream decoding architecture," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2007, pp. IV-401–IV-404.
- 17) G.Sumithra<sup>1</sup>, A.K.Devika, "Performance Analysis of Speaker Identification System Using GMM with VQ", *IJCNS*, Vol 4. No 1. Jan-Mar 2012 ISSN: 0975-8283.
- 18) V. Ruonala et al., "Dynamic tension EMG to characterize the effects of DBS treatment of advanced Parkinson's disease," *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL*, pp. 3248-3251, 2014.