International Journal of Scientific Research and Management (IJSRM)
||Volume||5||Issue||11||Pages||7474-7484||2017||
Website: www.ijsrm.in ISSN (e): 2321-3418
Index Copernicus value (2015): 57.47 DOI: 10.18535/ijsrm/v5i11.17

Analysis on Multi-Aggregative Factor Using Electromyographic Data

[1]P.Tamil Selvan, [2]Dr. Senthil Kumar A.V

[1]Research Scholar, Hindusthan College Of Arts & Science
[2]Research Supervisor Hindusthan College Of Arts & Science

Abstract:-
As in sample by sample classification, when a classifier is asked to combine information across multiple samples drawn from the same data source, the results are combined using a strategy such as majority vote. To solve the problem of classification failure (i.e., a hazard function) in multi sample classification, Multi-aggregative factored K-NN Classifier is proposed. This method evaluates the classification of multi sample 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 vote 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.

Keywords-[EMG, Motor Unit Action Potential, Additional Feature Sets, Classifier]

1. Introduction

Data mining is the process of revealing nontrivial, previously unknown and potentially useful information from large databases [1]. Discovering useful patterns hidden in a database plays an essential role in several data mining tasks, such as frequent pattern mining, weighted frequent pattern mining, and high utility pattern mining. Among them, frequent pattern mining is a fundamental research topic that has been applied to different kinds of databases, such as transactional databases, streaming databases and time series databases and various application domains, such as bioinformatics, Web click-stream and mobile environments. Nevertheless, relative importance of each item is not considered in frequent pattern mining. To address this problem, weighted association rule mining was proposed. In this framework, weights of items, such as unit profits of items in transaction databases, are considered. With this concept, even if some items appear infrequently, they might still be found if they have high weights [2-4]. However, in this framework, the quantities of items are not considered yet. Therefore, it cannot satisfy the requirements of users who are interested in discovering the itemsets with high sales profits, since the profits are composed of unit profits, i.e., weights, and purchased quantities. In view of this, utility mining emerges as an important topic in data mining field. Mining high utility itemsets from databases refers to finding the itemsets with high profits. Here, the meaning of itemset utility is interestingness, importance, or profitability of an item to users[5].

A pattern-based knowledge editing system for building clinical Decision Support Systems- Decision support in medicine is being more and more configured as an innovative and valuable way for promoting more consistent, effective, and efficient medical practices. In this respect, this paper proposes a pattern-based knowledge editing system to guide and assist the creation and formalization of condition-action clinical recommendations to be used in knowledge-based Decision Support Systems (in the following, DSSs) [6]. This system has been devised with the aim of: (i) offering a set of patterns for easily inserting and editing such clinical recommendations; (ii) synergistically combining multiple knowledge representation techniques to instantiate these patterns within hybrid knowledge bases (KBs), made of if–then rules built on the top of ontological vocabularies; (iii) reducing the complexity of the formalization process, by graphically guiding the definition of hybrid KBs that could be functional in the context of clinical DSSs and enabling their automatic encoding into machine executable languages[7]. A usability evaluation has been carried out, showing a good satisfaction of medical users with respect to the system implemented, and, thus, proving both the feasibility and usefulness of the approach proposed.

Nevertheless, it is extremely worth highlighting that the need of technical experts for editing and upgrading clinical recommendations in a knowledge-based DSS is a strong limitation for medical users [8]. As a matter of fact, actually, one prerequisite for the broad acceptance of such DSSs and their efficient application to medical settings is the guarantee of a high level of upgradability and maintainability, (i) to change clinical rules according to their evolution to implement medical progress in the treatment of individual diseases, or (ii) to adapt generic, site-independent clinical rules to a patient to be treated. Since updating the KB can require a continuous intervention, it is unthinkable that it cannot be done directly by doctors when needed. Also, by providing a direct access to the KB, doctors are encouraged to use clinical DSSs built on the top of it, since mostly entrusted with the suggestions generated starting from their expertise, especially if encoded by them.

In contrast to the intensive efforts made to develop knowledge based DSSs, the issue of providing solutions for easily editing and upgrading condition-action clinical rules into their KB has been widely neglected thus far. In this respect, this paper proposes a pattern-based knowledge editing system to guide and assist the creation and formalization of condition-action clinical recommendations to be used in knowledge-based DSSs [9].

A self-referential outlier detection method for quantitative motor unit action potential analysis- Quantitative MUAP analysis is often based on outlier detection, in the case of neurogenic conditions, the finding of MUAPs that are larger than the limit determined from a reference normal population. Such reference data is available from only a few sources and for only a few muscles. It would be desirable if muscles could serve as their own controls. The Henneman size principle determines the order of recruitment, following an exponential distribution of twitch force, motor neurone, motor unit, and MUAP size. Therefore, an outlier could be detected by being too large for the level of recruitment, as judged by its size

relative to the other MUAPs. This would improve the sensitivity of detecting neurogenic muscles and obviate the need for external reference data [10].

In the neurogenic IPs, all the absolutely large MUAPs should be detected but possibly also some MUAPs that are only relatively large. The results for MUAP Size Index and energy would be compared – they might prove to have complementary diagnostic value. The sensitivity and specificity of this technique can then be checked on larger groups of normal subjects and on subjects with known neurogenic conditions, although such subjects are hardly a gold standard. An ideal gold standard would be a muscle biopsy of an area studied by EMG. There is no certainty that a muscle or muscle region studied by EMG in a patient with a condition that could cause motor axonal injury and collateral re-innervation (e.g. polyneuropathy) is actually pathologically so-affected [11].

Bayesian aggregation versus majority vote in the characterization of non-specific arm pain based on quantitative needle electromyography- Methods for the calculation and application of quantitative electromyographic (EMG) statistics for the characterization of EMG data detected from forearm muscles of individuals with and without pain associated with repetitive strain injury are presented.

A classification procedure using a multi-stage application of Bayesian inference is presented that characterizes a set of motor unit potentials acquired using needle electromyography. The utility of this technique in characterizing EMG data obtained from both normal individuals and those presenting with symptoms of "nonspecific arm pain" is explored and validated. The efficacy of the Bayesian technique is compared with simple voting methods [12].

Quantitative electromyographic (EMG) data can be used to obtain reproducible and robust characterizations of the signature signal structures obtained from individual motor units (MUs). Through signal decomposition techniques applied to a needle-detected EMG signal, it is possible to observe the repeated occurrence of motor-unit potentials (MUPs) from the pool of motor units active during a given muscle contraction. The series of such potentials is referred to as a motor-unit potential train, or MUPT; these data may be used to characterize both the average shape of a MUP as well as to estimate the firing pattern of its generating MU [13].

In addition, by combining data simultaneously acquired using surface and needle electrodes, it is possible to correlate the data from these sources and obtain an estimate of the surface representation of the MUP (called an SMUP template) related to each MUPT. The SMUP is determined by using the firing times of the main spike of each individual MUP firing within a MUPT and relating these to the potential observed at a surface electrode overlying the needle uptake volume [14].

Clinical Decision Support and Malpractice Risk- Clinical Decision Support (CDS) refer to electronic technology used to enhance clinical decision making. For example, computerized physician order entry with integrated CDS in principle offers an electronic layer of review for ordering prescriptions [15].

An important feature of CDS is automated warnings issued whenever potential drug interactions or other contraindications arise. In practice, however, CDS systems often have been over inclusive in the

warnings they generate, to a point at which physician "alert fatigue" may in large part undermine the utility the systems offer. The current generation of CDS systems includes alert parameters for thousands of drug interaction types. Meanwhile, a recent review of empirical studies on computerized physician order entry with integrated CDS observed that physicians override automated warnings a substantial fraction of the time according to one study, in as many as 19 out of 20 instances. One paradoxical result of overly abundant warnings may be to exacerbate malpractice risk for physicians who either ignore or turn off CDS alerts, even as CDS systems create an audit trail to show that those physicians have done so [16-17].

2. Existing System

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. 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.

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

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.

List of Phases

1. Samples Collection and Preprocessing

2. Feature Selection
3. Samples Similarity Construction
4. Disease Prediction or Classification
5. Survival Prediction

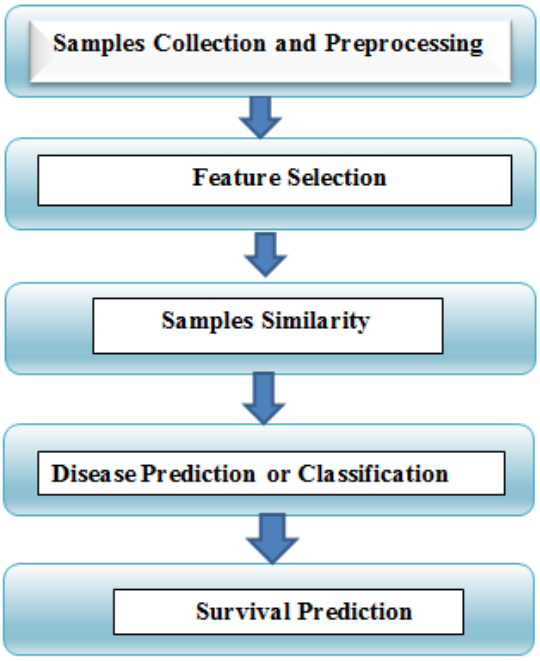

Figure 1: Research Architecture Flow

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, 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 … 15 is defined on each view of the data. 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 neighbors of the unknown

instance, which are found using this distance function, serve 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 neighbors 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.

Let (X_i, C_i) where $i = 1, 2\ldots\ldots, n$ 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'
$C_i \in \{1, 2, 3, \ldots\ldots, c\}$ 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 "$d(x, x_i)$" $i = 1, 2, \ldots\ldots,$ **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 $\mathbf{k}_i$ denotes the number of points belonging to the i^{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.*

Multi-aggregative factored KNN Algorithm

4. Experimental Results and Discussions

The most often metric used to determine the performance of classifier is accuracy. Since the accuracy is inappropriate when data is imbalanced, we used another metrics to compare the performance. The standard technique for evaluating classifier on imbalanced class is Receiver Operating Characteristic.

It shows KNN has constant accuracy even though the data has been randomized 30 times. Random Forest Tree can classify the result better than other classifiers. Recall measures how often a positive class instance in the dataset was predicted as a positive class instance by the classifier. Precision measure how often an instance that was predicted as positive that is actually positive.

Here we compare proposed and existing algorithms are

1. Naive Bayes 2. Decision Tree 3. KNN

Survival Probability: In this section, we evaluate memory usage for each algorithm with the same datasets as the runtime tests. Our algorithm, it guarantees Survival Probability as good as that of the state-of-the-art algorithm. Moreover, our algorithm presents the most outstanding results in many cases.

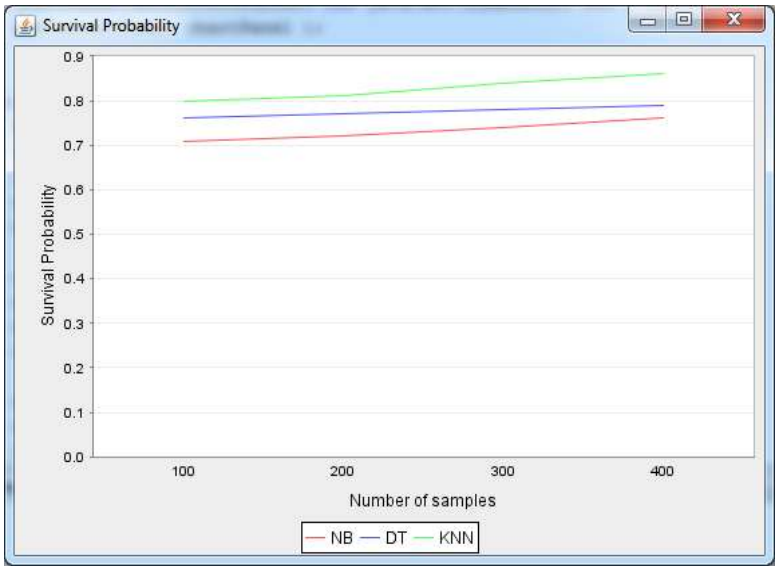

Figure 4.1-Survival Probability

No of Samples	Naive Bayes	Decision Tree	KNN
100	**0.71**	**0.76**	**0.80**
200	**0.72**	**0.77**	**0.81**
300	**0.74**	**0.78**	**0.84**
400	**0.76**	**0.79**	**0.86**

Table 4.1 Survival Probability Results

Accuracy (%): We can observe that our proposed outperforms the others in almost all of the cases. Our proposed linear structure to its trees instead of the previous tree form in order to minimize access times to search nodes. As a result, its advantages have a positive effect on reducing runtime in whole experiments. Especially as the minimum support threshold becomes lower, the difference of runtime between our algorithm and the others is bigger.

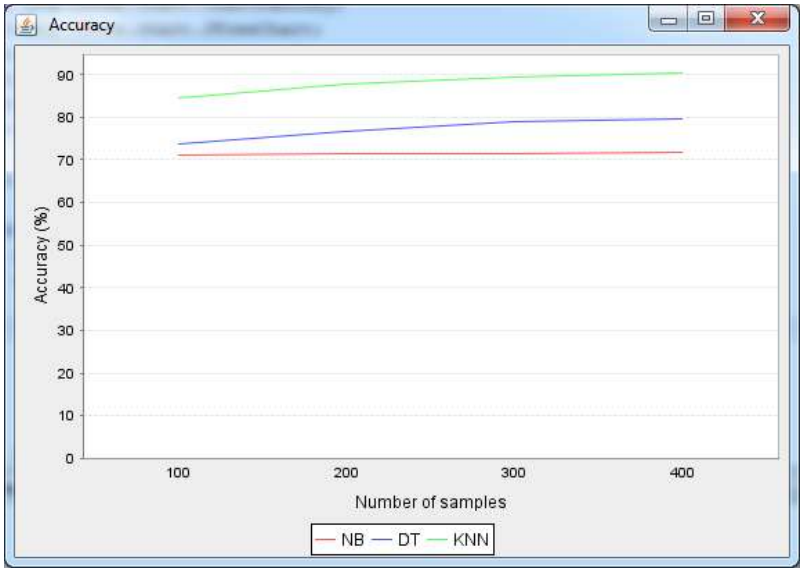

Figure4.2- Accuracy

No of Samples	Naive Bayes	Decision Tree	KNN
100	71.0	73.8	84.6
200	71.3	76.6	87.8
300	71.5	78.9	89.5
400	71.8	79.5	90.3

Table 4.2 Accuracy Results

Precision (%): Proposed algorithm shows the best Precision while the others have relatively poor performance, which indicates that our scheme can store these increasing attributes more efficiently than the other structures of the competitor algorithms. Through the above experimental results, we know that the proposed algorithm outperforms the others with respect to increasing transactions and items in terms of scalability as well as runtime and memory usage for the real datasets.

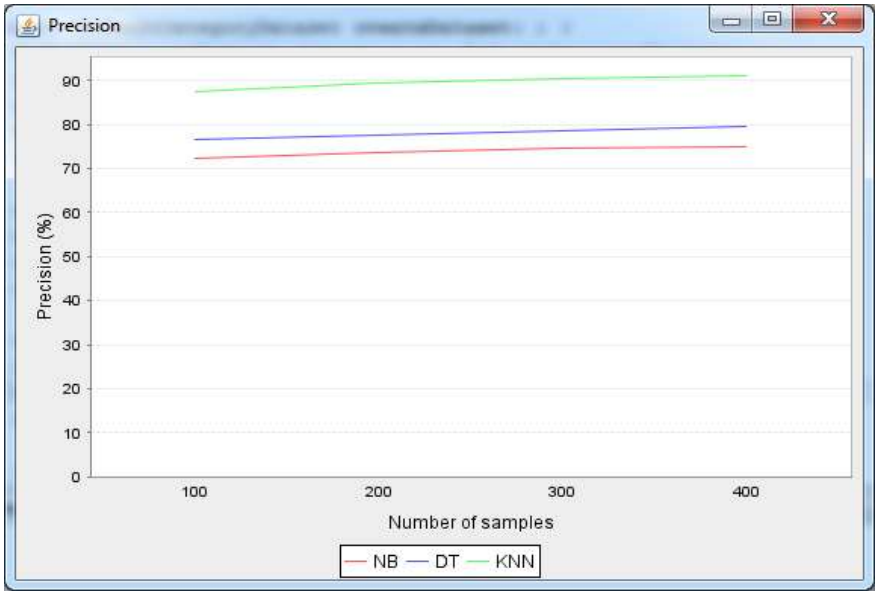

Figure 4. 3 –Precision

No of Samples	Naive Bayes	Decision Tree	KNN
100	72.5	76.7	87.5
200	73.5	77.5	89.4
300	74.5	78.7	90.5
400	75.0	79.5	91.0

Table 4.3 Precision Results

Recall(%): Through the above experimental results, we know that the proposed algorithm outperforms the others with respect to increasing transactions and items in terms of scalability as well as runtime and memory usage for the real datasets.

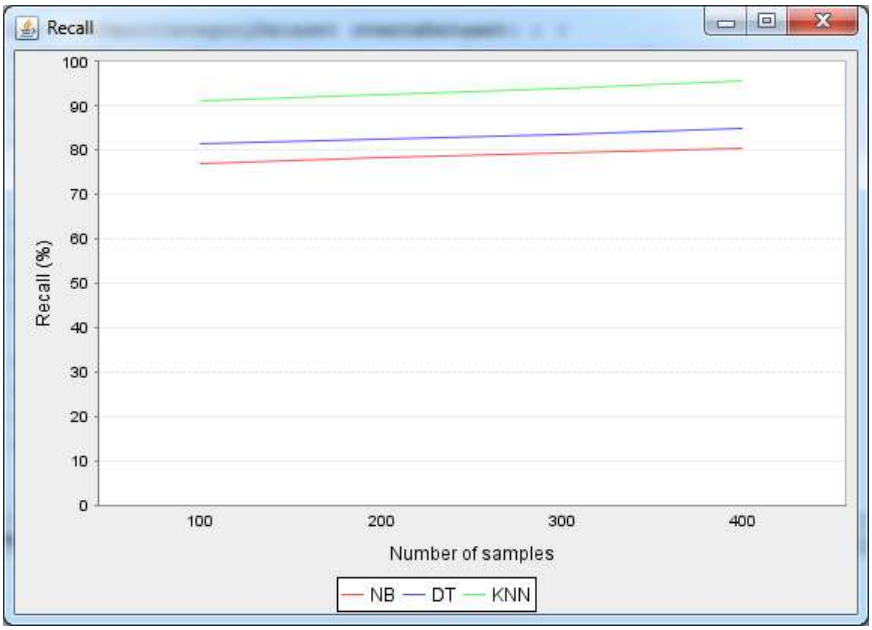

Figure 4.4 Recall

No of Samples	Naive Bayes	Decision Tree	KNN
100	77.0	81.5	91.0
200	78.5	82.5	92.4
300	79.5	83.5	93.9
400	80.5	84.8	95.5

Table 4.4 Recall Results

Our experimental results showed that KNN outstanding performance in terms of accuracy, precision, recall, memory usage, and scalability. Our proposed work significant improvement in classification accuracy, precision, recall and in survival probability. This significant effect indicates that use of a KNN in decision support systems has multiple benefits, both in terms of system accuracy and in terms of system transparency.

5. Conclusion

K-NN classification classifies instances based on their similarity to instances in the training data. As in sample by sample classification, when a classifier is asked to combine information across multiple samples drawn from the same data source, the results are combined using a strategy such as majority vote. To solve the problem of classification failure (i.e., a hazard function) in multisample classification, Multi-aggregative factored K-NN Classifier is proposed. This method evaluates the classification of multisample 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 vote 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. Our result shows that the K-Nearest Neighbor classifier is transparent, consistent, straightforward, simple to understand, high tendency to possess desirable qualities and easy to implement than most other machine learning techniques specifically when there is little or no prior knowledge about data distribution.

Reference

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. C. Christodoulou and C. Pattichis, "Medical diagnostic systems using ensembles of neural SOFM classifiers," in Proc. 6th IEEE Int. Conf. Electron., Circuits Syst., 1999, vol. 1, pp. 121–124.
3. L. I. Kuncheva,Combining Pattern Classifiers: Methods and Algorithms. New York, NY, USA: Wiley, 2004.
4. L. Rokach, "Ensemble-based classifiers,"Artif. Intell. Rev., vol. 33, no. 1/2, pp. 1–39, Feb. 2010.
5. D. W. Stashuk, "Decomposition and quantitative analysis of clinical electromyographic signals," Med. Eng. Phys., vol. 21, no. 6/7, pp. 389–404, Feb. 1999.
6. D. W. Stashuk and W. F. Brown, "Quantitative electromyography," in Neuromuscular Function and Disease,vol.1,W.F.Brown,C.F.Bolton, and M. J. Aminoff, Eds. Philadelphia, PA, USA: Saunders, 2002, pp. 311–348.
7. D. W. Stashuk and T. J. Doherty, "The normal motor unit action potential," in Neuromuscular Function and Disease, vol.1,W.F. Brown, C. F. Bolton, and M. J. Aminoff, Eds. Philadelphia, PA, USA: Saunders, 2002.
8. D. W. Stashuk, L. Pino, A. Hamilton Wright, T. Doherty, and S. Boe,Interpretation of QEMG data," presented at the General Meet. Amer. Assoc. Neuromuscular Electrodiagnostic Med.. Phoenix, AZ, USA, 2007.

9. J. V. Basmajian and C. J. De Luca, Muscles Alive: Their Functions Revealed by Electromyography, 5th ed. Baltimore, MD, USA: William &Wilkins, 1985.
10. R. L. Teach and E. H. Shortliffe, "An analysis of physician attitudes regarding computer-based clinical consultation systems,"Comput. Biomed. Res., vol. 14, no. 6, pp. 542–558, Dec. 1981.
11. A. X. Garg, N. K. J. Adhikari, H. McDonald, M. P. Rosas-Arellano,P. J. Devereaux, J. Beyene, J. Sam, and R. B. Haynes, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review,"J. Amer. Med. Assoc., vol. 293, no. 10, pp. 1223–1238, 2005.
12. M. Greenberg and M. S. Ridgely, "Clinical decision support and malpractice risk,"J. Amer. Med. Assoc., vol. 306, no. 1, pp. 90–91, 2011.
13. A. Dupuy and R. M. Simon, "Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting,"J. Nature Cancer Inst., vol. 99, no. 2, pp. 147–157, Nov. 2007.
14. Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos. (2010, May). Highdimensional pattern regression using machine learning: From medical images to continuous clinical variables. NeuroImage[Online]. 50(4), pp. 1519-1535. Available:http://www.sciencedirect.com/science/article/pii/S1053811909013810
15. H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting,"IEEE Trans. Bio-Med. Eng., vol. 57, no. 12, pp. 2936–2946, Dec. 2010.
16. G. Pfeiffer, "The diagnostic power of motor unit potential analysis: An objective Bayesian approach,"Muscle Nerve, vol. 22, no. 5, pp. 584–591, 1999.
17. Dr.Senthil kumar A.V, and P.Tamilselvan "Multisample Classification In Clinical Decisions Using Multi-Aggregative Factored K-NN Classifier," Internal Journal of Trend scientific research and development ,vol. 01, no. 6, pp. 2456-6470, Oct-2017.