



- (51) International Patent Classification:
G06F 17/00 (2006.01) *G06F 17/10* (2006.01)
G06F 15/18 (2006.01)
- (21) International Application Number:
PCT/CA2011/001085
- (22) International Filing Date:
4 October 2011 (04.10.2011)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/439,413 4 February 2011 (04.02.2011) US
- (71) Applicant (for all designated States except US): **HOLLAND BLOORVIEW KIDS REHABILITATION**

HOSPITAL [CA/CA]; 150 Kilgour Road, Toronto, Ontario M4G 1R8 (CA).

- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **NIKJOO SOUKHTABANDANI, Mohammad** [IR/CA]; Apt #1317-15 Northtown Way, Toronto, Ontario M2N 7A2 (CA). **CHAU, Thomas T. K.** [CA/CA]; 77 Hogarth Avenue, Toronto, Ontario M4K 1K2 (CA).
- (74) Agent: **GASTLE AND ASSOCIATES**; 12 Bridge St. Suite 202, Lakefield, Ontario K0L 2H0 (CA).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

[Continued on next page]

(54) Title: REPUTATION-BASED CLASSIFIER, CLASSIFICATION SYSTEM AND METHOD

(57) Abstract: Disclosed herein is a reputation-based classification system, method and device. In one embodiment, classification is implemented using two or more classifiers, wherein each classifier is trained using a training data set; a respective overall performance is measured for each trained classifier using a validation data set; a respective reputation value is assigned to each of the trained classifiers representative of the measured respective overall performance thereof; and the test data is classified via combination of the two or more trained classifiers as a function of their respective reputation values.

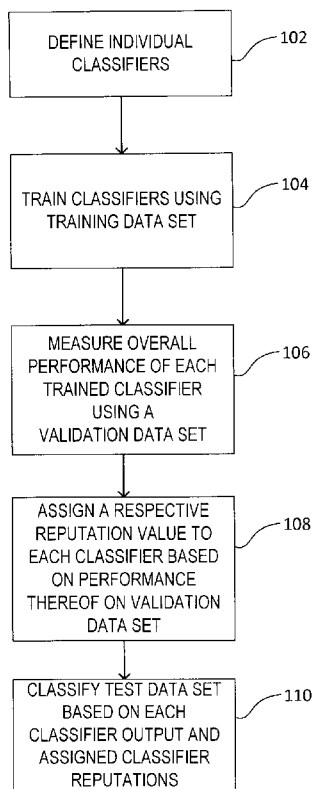


FIGURE 1





HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,

Published:

— with international search report (Art. 21(3))

REPUTATION-BASED CLASSIFIER, CLASSIFICATION SYSTEM AND METHOD

FIELD OF THE DISCLOSURE

[0001] The present disclosure relates to classifiers, and in particular, to a reputation-based classifier, classification system and method.

5

BACKGROUND

[0002] The exercise of combining classifiers is primarily driven by the desire to enhance the performance of a classification system. There may also be problem-specific rationale for integrating several individual classifiers. For example, a designer may have access to different types of features from the same study participant. For instance, in the human identification problem, the participant's voice, face, and handwriting provide different types of features. In such instances, it may be sensible to train one classifier on each type of feature. In other situations, there may be multiple training sets, collected at different times or under slightly different circumstances. Individual classifiers can be trained on each available data set. Lastly, the demand for classification speed in online applications may necessitate the use of multiple classifiers.

[0003] Traditionally, the goal of these methods is to improve classification accuracy by employing multiple classifiers to address the complexity and non-uniformity of class boundaries in the feature space. For example, classifiers with different parameter choices and architectures may be combined so that each classifier focuses on the subset of the feature space where it performs best. Well-known examples of these methods include bagging and boosting. Given the universal approximation ability of neural networks such as multilayer perceptrons and radial basis functions, there is theoretical appeal to combine several neural network classifiers to enhance classification. Indeed, several methods have been developed for this purpose, including, for example, optimal linear combinations and mixture of experts, and negative correlation and evolving neural network ensembles. In these methods, all base classifiers are generally trained on the same feature space (either using the entire training set or subsets of the training set). While these methods have

proven effective in many applications, they are associated with numerical instabilities and high computational complexity in some cases.

[0004] Another approach to classifier combination is to train the base classifiers on different feature spaces. This approach is advantageous in combating the undesirable effects associated with high-dimensional feature spaces (curse of dimensionality). Moreover, the feature sets can be chosen to minimize the correlation between the individual base classifiers to further improve the overall accuracy and generalization power of classification. These methods are also highly desirable in situations where heterogeneous feature combinations are used.

[0005] Combination of classifiers based on different features has been generally accomplished through fixed classification rules. These rules may select one classifier output among all available outputs (for example, using the minimum or maximum operator), or they may provide a classification decision based on the collective outputs of all classifiers (for example, using the mean, median, or voting operators). Among the latter, the simplest and most widely applied rule is the majority vote. Many authors have demonstrated that classification performance improves beyond that of the single classifier scenario when multiple classifier decisions are combined via a simple majority vote.

[0006] The following notation is used for the purpose of illustrating the application of, and inherent deficiencies in, the majority voting approach. Assume the time series, S , is the pre-processed version of an acquired signal. Also let $\Theta = \{\theta_1, \theta_2, \dots, \theta_L\}$ be a set of $L \geq 2$ classifiers and $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ be a set of $c \geq 2$ class labels, where $\omega_j \neq \omega_k$, $\forall j \neq k$. Without loss of generality, $\Omega \subset \mathbb{N}$. The input of each classifier is the feature vector $x \in \mathfrak{R}^{n_i}$, where n_i is the dimension of the feature space for the i^{th} classifier θ_i , whose output is a class label ω_j , $j = 1, \dots, c$. In other words, the i^{th} classifier, $i = 1, \dots, L$, is a functional mapping, $\mathfrak{R}^{n_i} \rightarrow \Omega$, which for each input x gives an output $\theta_i(x) \in \Omega$. Generally, the classifier function could be linear or non-linear. It is assumed that for the i^{th} classifier, a total number of d_i subjects are assigned for training. The main goal of

combining the decisions of different classifiers is to increase the accuracy of the class selection.

[0007] In a multi-classifier system, the problem is to arrive at a global decision $\Theta^*(x) = \omega_j$ given a number of local decisions $\theta_i(x) \in \Omega$, where generally:

$$5 \quad \theta_1(x) = \theta_2(x) = \dots = \theta_L(x).$$

In the literature, a classical approach for solving this problem is *majority voting*. To express this idea mathematically, we define an indicator function:

$$I_i(x, w_j) = I(\theta_i(x) = w_j) = \begin{cases} 1, & \text{when } \theta_i(x) = w_j, \\ 0, & \text{otherwise.} \end{cases}$$

from which the majority voting rule can be expressed as follows:

$$10 \quad \Theta^*(x) = \begin{cases} \omega_{\max}, & \text{if } \max_{\omega_j} I_{\Omega}(x, \omega_j) > L/2, \\ Q, & \text{otherwise,} \end{cases}$$

where $\omega_{\max} = \arg \max_{\omega_j} I_{\Omega}(x, \omega_j)$, $I_{\Omega}(x, \omega_j) = \sum_{i=1}^L I_i(x, \omega_j)$, $j = 1, \dots, c$, and $Q \notin \Omega$ is the rejection state. In other words, given a feature vector, each classifier votes for a specific class. The class with the majority of votes is selected as the *candidate class*. If the candidate class earns more than half of the total votes, it is selected as the final output of the system. Otherwise, the feature vector is rejected by the system.

[0008] The majority voting algorithm is computationally inexpensive, simple to implement and applicable to a wide array of classification problems. Despite its simplicity, majority voting can significantly improve upon the classification accuracies of individual classifiers. However, this method suffers from a major drawback; the decision heuristic is strictly democratic, meaning that the votes from different classifiers are always equally weighted, regardless of each the performance of individual classifiers. Therefore, votes of weak classifiers, i.e., classifiers whose performance only slightly exceeds that of the random classifier, can diminish the overall performance of the system

when they have the majority. To exemplify this issue, consider a classification system with $c = 2$ classes, $\Omega = \{\omega_1, \omega_2\}$, and $L = 3$ classifiers, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, where two are weak classifiers with 51% average accuracy while the remaining one is a strong classifier with 99% average accuracy. Now assume that for a specific feature vector both the weak classifiers vote for ω_1 but the strong classifier votes for ω_2 . Based on the majority voting rule, ω_1 is preferred over ω_2 , which is most likely an incorrect classification.

[0009] While the notion of weighted majority voting was further introduced to improve performance by incorporating classifier-specific beliefs which reflect each classifier's uncertainty about a given test case, this improvement still suffers from numerous drawbacks. For example, this method, for example as proposed by Xu et al. in Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition (IEEE Transactions on Systems, Man, and Cybernetics, Vol. 22, No. 3, May/June 1992) does not deal with the degenerate case when one or more beliefs are zero - a situation likely to occur in multi-class classification problems. Moreover, as such classifiers generally rely on the training data set to derive beliefs values for each classifier, this approach, therefore, risks overfitting the classifiers to the training set and a consequent degradation in generalization power.

[0010] Therefore, there remains a need for a classifier, classification system and method that overcome at least some of the drawbacks of known techniques, or at least, provides a useful alternative.

[0011] This background information is provided to reveal information believed by the applicant to be of possible relevance to the invention. No admission is necessarily intended, nor should be construed, that any of the preceding information constitutes prior art against the invention.

SUMMARY

[0012] An object of the invention is to provide a reputation-based classifier, classification system and method that overcome at least some of the drawbacks of known

techniques, or at least, provides a useful alternative thereto. In accordance with one embodiment of the invention, there is provided a method for classifying test data using two or more classifiers, the method comprising the steps of: training said two or more classifiers using a training data set; measuring a respective overall performance of each of said trained classifiers using a validation data set; assigning a respective reputation value to each of said trained classifiers representative of said respective overall performance thereof; and classifying the test data via combination of said two or more trained classifiers as a function of said respective reputation values.

[0013] In accordance with another embodiment, there is provided a method for classifying test data using two or more classifiers, the method comprising the steps of: classifying the test data using each of said two or more classifiers to obtain respective classifications therefrom; calculating a highest likelihood classification for the test data as a function of said respective classifications and as a function of a respective overall performance value previously measured for each of the two or more classifiers; and outputting said highest likelihood classification as global classification for the test data.

[0014] In accordance with another embodiment of the invention, there is provided a computer-readable medium having statements and instructions stored thereon that, upon execution by a processor, automatically implement the steps of the above methods.

[0015] In accordance with another embodiment of the invention, there is provided a computer-readable medium having statements and instructions stored thereon for execution by a processor of a computing device in automatically classifying input test data, the statements and instructions comprising: two or more encoded classifiers each configured to output respective local data classifications; a training module for training said two or more classifiers on a training data set; a validation module for measuring a respective overall performance value for each of said trained classifiers using a validation data set, and assigning a respective reputation value to each of said trained classifiers as a function thereof; and a classification module for locally classifying the test data via each of said two or more trained classifiers, and globally classifying the test data as a function

of each said respective reputation value and said respective local data classifications output from said trained classifiers on the test data.

[0016] In accordance with another embodiment of the invention, there is provided a computer-readable medium having statements and instructions stored thereon for execution by a processor of a computing device in automatically classifying input test data, the statements and instructions comprising: two or more encoded classifiers each trained to output respective local data classifications; a respective reputation value assigned to each of said two or more classifiers representative of a respective overall performance thereof; and a classification module for locally classifying the test data via each of said two or more trained classifiers, and globally classifying the test data as a function of each said respective reputation value and said respective local data classifications output from said trained classifiers on the test data.

[0017] In accordance with another embodiment of the invention, there is provided a device for classifying test data using two or more classifiers, the device comprising: a processor; an input for receiving test data to be classified; an output for outputting a global classification of the test data; a computer-readable data storage device operatively coupled to said processor, input and output, and having stored thereon statements and instructions for execution by said processor in classifying the test data, said statements and instructions comprising: two or more encoded classifiers each trained to output respective local data classifications; a respective reputation value assigned to each of said two or more classifiers representative of a respective overall performance thereof; and a classification module for locally classifying the test data via each of said two or more trained classifiers, globally classifying the test data as a function of each said respective reputation value and said respective local data classifications output from said trained classifiers on the test data, and communicating a resulting global classification to said output.

[0018] Other aims, objects, advantages and features of the invention will become more apparent upon reading of the following non-restrictive description of specific

embodiments thereof, given by way of example only with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE FIGURES

5 [0019] Several embodiments of the present disclosure will be provided, by way of examples only, with reference to the appended drawings, wherein:

[0020] Figure 1 is a high level flow chart of a reputation-based classification method, in accordance with one embodiment of the invention;

[0021] Figure 2 is a flow chart of an exemplary reputation-based classification method, in accordance with one embodiment of the invention;

10 [0022] Figure 3 is a schematic diagram of a reputation-based classification device, in accordance with one embodiment of the invention;

[0023] Figure 4 is a schematic diagram of an experimental setup for validating use of a reputation-based classification method for classifying dual-axis cervical accelerometry signals as representative of healthy or unhealthy swallowing events, in accordance with
15 one embodiment of the invention;

[0024] Figure 5 is an exemplary graphical representation of dual-axis cervical accelerometry data for a healthy swallowing event, in accordance with one embodiment of the invention;

[0025] Figure 6 is an exemplary graphical representation of dual-axis cervical
20 accelerometry data for an unhealthy swallowing event, in accordance with one embodiment of the invention;

[0026] Figure 7 is a graphical representation of the sensitivity, specificity and accuracy of single-axis and dual-axis accelerometry classifiers, in accordance with one embodiment of the invention;

[0027] Figure 8 is a parallel axes plot depicting internal representation of safe and unsafe swallows acquired by a reputation-based classifier, in accordance with one embodiment of the invention; and

[0028] Figure 9 is a graphical performance comparison between results of a traditional classifier combination method and that of a reputation-based classifier implemented in accordance with one embodiment of the invention.

DETAILED DESCRIPTION

[0029] It should be understood that the disclosure is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The disclosure is capable of other embodiments and of being practiced or of being carried out in various ways. Also, it is to be understood that the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of "including," "comprising," or "having" and variations thereof herein is meant to encompass the items listed thereafter and equivalents thereof as well as additional items. Unless limited otherwise, the terms "connected," "coupled," and "mounted," and variations thereof herein are used broadly and encompass direct and indirect connections, couplings, and mountings. In addition, the terms "connected" and "coupled" and variations thereof are not restricted to physical or mechanical or electrical connections or couplings. Furthermore, and as described in subsequent paragraphs, the specific mechanical or electrical configurations illustrated in the drawings are intended to exemplify embodiments of the disclosure. However, other alternative mechanical or electrical configurations are possible which are considered to be within the teachings of the instant disclosure. Furthermore, unless otherwise indicated, the term "or" is to be considered inclusive.

[0030] As introduced above, traditional majority voting classification systems are often inadequate in accurately classifying data/signals, as can weighted majority voting classifications (*e.g.* based on preset classifier beliefs), for the reasons outlined above and other reasons that will be readily appreciated by the skilled artisan. Accordingly, and as discussed in greater detail below, the embodiments of the invention described herein

provide an alternative to current classification systems, whereby the past performance of at least some classifiers is taken into account in arriving at a final (i.e. global) classification result by way of respective classifier reputations.

[0031] For instance, the general disregard for past classifier performance in available classification systems allows for such systems to routinely rely on weak classifiers, which can ultimately result in inaccurately classifying new results, thus diminishing the overall performance of the system. Accordingly, and in accordance with one embodiment of the invention, the classifier and classification methods considered herein mitigate the risk that the overall decision computed thereby will be unduly influenced by poorly performing classifiers by assigning different reputations to the decisions of different classifiers based on their past performances, thus allowing the classifier and classification methods considered herein to account for respective classifier performances, and thus their respective reliability, in computing a new overall decision. For example, in one embodiment, a reputation value is first calculated for each classifier using a known (e.g. validation) data set. Such reputation values may, in accordance with different embodiments, be calculated as a function of a measured overall performance of each classifier, for example, on the known data set. Namely, the overall performance of a given classifier may be defined, in some embodiments, as the overall accuracy (e.g. percentage of correct classifications) of this classifier in classifying the known data set, which accuracy can then be used in evaluating the likelihood that this classifier's output on an unknown (e.g. test) data set is accurate or not. Accordingly, based on these respective reputation values, an effective weight can be associated with each classifier, to be accounted for in subsequent classifications. As will be described in greater detail below, execution of a reputation-based classifier, as considered herein, can increase the overall performance of such systems.

[0032] Reputation typically refers to the quality or integrity of an individual component within a system of interacting components. In accordance with the embodiments of the invention described herein, the concept of reputation is applied to judiciously combine local decisions of multiple classifiers for the purpose of globally classifying, for example, various signals, such that upon classification, a reasonable,

accurate and/or informative differentiation between such signals is achieved. In the examples provided below, reputation-based classification is used to differentiate between safe and unsafe swallows in aspiration detection, however, it will be appreciated that the below-described principles are readily applicable to other types of data/signals, such as in the classification of other physiological and/or biomechanical signals, which may, in some embodiments, allow or facilitate differentiation of such signals in providing/developing access technologies for candidates with serious disabilities, in controlling prosthetics (e.g. classification of muscle contractions measured as MMG and/or EMG), and the like. In other embodiments, the use of reputation-based classification may rather be used for security purposes, for example in monitoring human activity in categorizing such activity as safe, suspicious or dangerous, for example. These and other such exemplary applications will be readily apparent to the person of ordinary skill in the art upon reference to the following description, and therefore, should not be considered to depart from the general scope and nature of the present disclosure. Namely, the embodiments of the invention herein described may be readily applied to different types of data/signals, wherein the interchangeable and liberal use herein of such terms as data, signal, data set, etc. should not be construed as limiting to the general scope of the present disclosure.

[0033] It will be further understood that while the various embodiments of classifiers, classification systems and methods are described below in general terms, such embodiments can be readily encompassed within and/or encoded and configured for implementation by a computing device or the like, which device for example encompassing one or more processors operatively coupled to one or more computer-readable media having encoded therein statements and instructions that, when implemented by the processor(s), implement the classifiers, classification systems and methods considered herein. Such computing devices may include, but are not limited to, multiple purpose computers such as desktops, laptops, palmtops and the like, dedicated computing devices and/or platforms such as for example, biomedical and/or biomechanical devices, diagnostic devices, monitoring devices and/or other such application-specific devices, and/or other types of dedicated, centralized, distributed

and/or networked computing device/platform. Examples of such devices will be described in greater detail below.

[0034] As introduced above, the general principle considered herein is to differentially weigh classifier decisions on the basis of their past performance. Namely, this novel fusion approach extends from the majority voting concept to acknowledge the past performance of classifiers, thus mitigating the risk of the overall decision being unduly influenced by poorly performing classifiers. Illustratively and following from the above-introduced notation with respect to majority voting approaches, and in accordance with one embodiment of the invention as schematically illustrated in Figure 1, the past performance of the i^{th} classifier in a reputation-based classifier can be defined as,

$$r_i \in \mathbb{R}, 0 \leq r_i \leq 1,$$

wherein 1 signifies a strong classifier (high accuracy) and 0 denotes a weak classifier. For each feature vector, both the majority vote and the reputation of each classifier contribute to the final global decision. The collection of reputation values for L classifiers constitutes the reputation set $R = \{R_1, R_2, \dots, R_L\}$. Each classifier is mapped to a real-valued reputation, r_i , namely

$$r_i = r(\theta_i) = \alpha, i = 1, \dots, L,$$

where

$$r : \Theta \rightarrow [0, 1], \alpha \in \mathbb{R} \text{ and } 0 \leq \alpha \leq 1.$$

[0035] To determine the reputation of each classifier, a validation set is utilized in addition to the classical training set. Specifically, in one embodiment, the overall (e.g. class-independent) performance of the trained classifiers on the validation data determines their reputation values. Namely, a given reputation value may be assigned to a given classifier as a function of an overall performance thereof on the validation set, for example determined as an accuracy or percentage of correct classifications achieved by this given classifier over a known data set, i.e. a labeled data set.

[0036] The following provides one illustrative embodiment of the reputation-based classification considered herein, as schematically depicted in Figure 1.

[0037] For a classification problem with $c \geq 2$ classes, $L \geq 2$ individual classifiers are designed and developed (i.e. step 102 of Figure 1). In one embodiment, the individual classifiers are independent, namely by using different training sets or using various resampling techniques such as bagging and boosting, for example. In general, there are no restrictions on the number of classifiers L and this value can be either an odd or an even number. Also, it should be noted here that, in general, the feature space dimension, n_i , of each classifier could be different and the number of training exemplars, d_i , for each classifier could be unique.

[0038] After training the L classifiers individually (step 104), the respective performance of each classifier is evaluated using the validation set (step 106) and a reputation value is assigned to each classifier (step 108). The validation sets are generally disjoint from the training sets; however, in one embodiment, it will be appreciated that the validation set may comprise or consist of the entire training set, or a subset thereof. It is important to note that here two different types of data sets are used, each with its own purpose. The first one is the traditional training set which is used repeatedly until the classifier is satisfactorily trained. In contrast, the second set consists of a validation set used to calculate the reputation values of individual classifiers, which should not be confused with the weights occasionally applied in traditional weighted majority voting methods, which weights generally represent preset beliefs defined with respect to the classifiers during the training phase.

[0039] In one embodiment, the accuracy of each classifier is estimated with the corresponding validation set and normalized to $[0,1]$ to generate a reputation value. For instance, a classifier, θ_i , with 90% overall accuracy (e.g. wherein classifier θ_i accurately classifies 90% of the elements of the validation set mentioned above) has a reputation $r_i = 0.9$. As will be discussed in greater detail below, this classifier can then be assumed to have a relatively high likelihood of accurately classifying new data, and a relatively low likelihood of inaccurately classifying new data, which likelihood can now, in accordance

with this embodiment, be accounted for in evaluating the relative impact this classifier may have when combined with other classifiers in outputting a global classification for new data.

[0040] Once each of the system's classifiers have been trained and assigned a
 5 respective reputation value in accordance with its overall performance on the validation data set, the system may be utilized to classify new test subjects, that is test data sets to be classified in accordance with a fused global classification based on the trained classifiers and their respective reputation values (step 110).

[0041] The following describes one illustrative embodiment of a reputation-based
 10 classification system, schematically depicted in Figure 2.

[0042] For each feature vector, x , in the test set, L decisions are obtained using the L distinct classifiers (step 202 of Figure 2):

$$\Omega(x) = \{\theta_1(x), \theta_2(x), \dots, \theta_L(x)\}$$

[0043] To arrive at a final decision, and in accordance with one embodiment, the
 15 votes of the highest reputation value classifiers are first considered rather than simply selecting the majority class. In one embodiment, the reputation values of the classifiers are sorted in descending order,

$$R^* = \{r_{1^*}, r_{2^*}, \dots, r_{L^*}\}$$

such that $r_{1^*} \geq r_{2^*} \geq \dots \geq r_{L^*}$. Then, using this set, the classifiers are ranked to obtain a
 20 reputation-ordered set of classifiers, Θ^* :

$$\Theta^* = \begin{pmatrix} \theta_{1^*} \\ \theta_{2^*} \\ \vdots \\ \theta_{L^*} \end{pmatrix}$$

[0044] In one embodiment, respective votes (e.g. local classifier outputs) from a subset of the reputation-ordered set of classifiers having highest respective reputation values are first considered (step 204). The respective local outputs of the first m classifiers of the reputation-ordered set of classifiers may be compared (step 206), and, upon each local output of this subset coinciding with a same output (i.e. identical output class labels) – step 208, this same output may be selected as global classification for the test data (step 210).

[0045] In this embodiment, the votes of the first m elements of the reputation-ordered set of classifiers (step 204) are considered (step 206), with

$$m = \begin{cases} \frac{L}{2}, & \text{if } L \text{ is even,} \\ \frac{L+1}{2}, & \text{if } L \text{ is odd.} \end{cases}$$

[0046] If the top m classifiers vote for the same class, ω_j (step 208), the majority vote is accepted and ω_j is retained as the final decision of the system (step 210). However, if the votes of the first m classifiers are not equal, the classifiers' individual reputations are then taken into account (step 212).

[0047] Let $p(\omega_j)$ be the prior probability of class ω_j . As before,

$$\Theta(x) = \{\theta_1(x), \theta_2(x), \dots, \theta_L(x)\}$$

represents the local decisions made by different classifiers about the input vector x . The probability that the combined classifier decision is ω_j given the input vector x and the individual local classifier decisions is denoted as the posterior probability,

$$p(\omega_j | \theta_1(x), \theta_2(x), \dots, \theta_L(x))$$

[0048] In one embodiment, the class that maximizes this property is selected, as defined by:

$$\arg \max_{\omega_j} p(\omega_j | \theta_1(x), \theta_2(x), \dots, \theta_L(x)), \quad j = 1, \dots, c$$

[0049] To estimate the posterior probability, and in accordance with one embodiment, the Bayes formula is used, as defined below, wherein the argument x was dropped for simplicity.

$$p(\omega_j|\theta_1, \dots, \theta_L) = \frac{p(\theta_1, \dots, \theta_L|\omega_j)p(\omega_j)}{p(\theta_1, \dots, \theta_L)},$$

- 5 where $p(\theta_1, \dots, \theta_L|\omega_j)$ is the likelihood of ω_j and $p(\theta_1, \dots, \theta_L)$ is the *evidence factor*, which is estimated using the *law of total probability*

$$p(\theta_1, \dots, \theta_L) = \sum_{j=1}^c p(x, \theta_1, \dots, \theta_L|\omega_j)p(\omega_j).$$

By assuming that the classifiers are independent of each other, the likelihood can be rewritten as follows:

10
$$p(\theta_1, \dots, \theta_L|\omega_j) = p(\theta_1|\omega_j) \dots p(\theta_L|\omega_j).$$

thus finally obtaining:

$$p(\omega_j|\theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i|\omega_j)p(\omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i|\omega_t)p(\omega_t)}$$

- [0050] The local likelihood functions $p(\theta_i|\omega_j)$ can be estimated by the reputation values calculated above. When the correct class is ω_j and classifier θ_i classifies x into class ω_j , *i.e.*, $\theta_i(x)=\omega_j$, the following is taken as true:
- 15

$$p(\theta_i = \omega_j|\omega_j) = r_i$$

In other words, $p(\theta_i = \omega_j | \omega_j)$ is taken as the probability that the classifier θ_i correctly classifies x into class ω_j when x actually belongs to this class. In this embodiment, this probability is exactly equal to the reputation value of the classifier (e.g. as defined by the classifier's previously measured overall performance on the validation set). On the other hand, when the classifier categorizes x incorrectly, i.e., $\theta_i(x)$ is not equal to ω_j given that the correct class is ω_j , then the complement of the reputation value can be used (e.g. the previously measured percentage of incorrect classifications by the classifier on the validation set):

$$p(\theta_i \neq \omega_j | \omega_j) = 1 - r_i.$$

10 [0051] When there is no known priority among classes, equal prior probabilities can be assumed, hence,

$$p(\omega_1) = p(\omega_2) = \dots = p(\omega_c) = \frac{1}{c}$$

Thus, for each class, ω_j , a posteriori probability can be estimated as given by the above equations. The class with the highest posteriori probability, i.e. the highest likelihood classification, is thus selected as the final decision of the system (step 214) and the input subject x is categorized as belonging to this class.

[0052] As will be appreciated by the skilled artisan, the above allows for the overall performance of each classifier on the validation data set to be used in influencing the impact local classifications from these classifiers may have on the global classification of the system, while mitigating degenerate cases commonly encountered when implementing prior art methods, amongst other advantages. Namely, by assigning a respective reputation value to each classifier, the likelihood of a given class within the system may be calculated not only as a function of each local classifier output, but also as a function of a respective reputation value for each classifier who's output coincides with the given class (i.e. a respective likelihood that each of these classifiers are correct based on a previously measured overall performance thereof on the validation set), and as a

function of a respective reputation value complement for each classifier who's output does not coincide with this given class (i.e. a respective likelihood that each of these classifiers are incorrect based on a previously measured overall performance thereof on the validation set).

5 [0053] In one embodiment, the advantage of the reputation-based approach as considered herein over the majority voting approach lies in the fact that the former has a higher probability of correct consensus and a faster rate of convergence to the peak probability of correct classification.

[0054] Referring now to Figure 3, and in accordance with one embodiment of the
10 invention, a classification device, generally referred to using the numeral 300, will now be described. In this embodiment the device 300 comprises a processor 302, an input 304 for receiving test data 306 to be classified and an output 308 for outputting a global classification 310 of the test data 306. The device 300 further comprises a computer-readable data storage device 312 operatively coupled to the processor 302, input 304 and
15 output 308, and having stored thereon statements and instructions for execution by the processor 302 in classifying the test data 306. In this particular embodiment, the statements and instructions encoded on the storage device 312 comprise two or more encoded classifiers 314 each trained to output a respective local data classification for the test data 306, and respective reputation values (R-values 316) assigned to these classifiers
20 314 and representative of a respective overall performance thereof. A classification module 318 is also provided for locally classifying the test data via each of the two or more trained classifiers 314, globally classifying the test data as a function of each respective reputation value 316 and the respective local data classifications output from the trained classifiers 314, and communicating the resulting global classification 310 to
25 the output 308. In one embodiment, the classification module is configured to compute the global classification 310, upon execution by the processor, in accordance with the classification steps discussed above.

[0055] In one embodiment, the device 300 may further comprises an optional validation module 320 for measuring an overall performance of each encoded classifier

on a validation set 322 received at the input 304 in defining the respective reputation values 316. In yet another embodiment, the device 300 may further comprise an optional training module 324 for training each encoded classifier 314 on a training set 326 received at the input 304.

5 [0056] It will be appreciated by the skilled artisan that the above-described embodiment of a classification device may be implemented in various forms, including, but not limited to, a dedicated device or computing platform, a dedicated classification platform implemented on one or more local and/or distributed computing devices, and/or other such system architectures as will be readily apparent to the skilled artisan.
10 Furthermore, while specific modules and data units are identified distinctly within the above-described embodiment, it will be appreciated that such distinctiveness is provided herein solely for the purpose of providing a clear description of the various features and elements of the device, and that such features and elements may be integrated within a same module or platform, or again distributed over various modules or platforms to
15 achieve a similar effect. Such variations are thus intended to fall within the general scope and nature of the present disclosure.

[0057] Reference will now be made to the following non-limiting examples, in which some of the above-proposed approaches to signal classification are applied to the classification of physiological signals, in accordance with exemplary embodiments of the
20 invention.

EXAMPLE 1

[0058] The accelerometric measurement of swallowing activity has been suggested as a potential non-invasive tool to assist in day-to-day management of swallowing difficulties in neurogenic dysphagia. Various vibratory signal features and
25 complementary measurement modalities have been put forth in the literature for the potential discrimination between safe and unsafe swallowing. To date, automatic classification of swallowing accelerometry has exclusively involved a single-axis of vibration although a second axis is known to contain additional information about the nature of the swallow.

[0059] In the following example, a large corpus of dual-axis accelerometric signals were collected from older adults referred to videofluoroscopic examination on the suspicion of dysphagia. A reputation-based classifier combination was then invoked to automatically categorize the dual-axis accelerometric signals into safe and unsafe swallows, as labeled via videofluoroscopic review. With selected time, frequency and information theoretic features, the reputation-based algorithm distinguished between safe and unsafe swallowing with an accuracy of (80.48+/-5.0) and provided interesting insight into the accelerometric differences between the two classes of swallows. Given its computational efficiency, reputation-based classification of dual-axis accelerometry provides, in accordance with one embodiment, a viable option for point-of care swallow assessment where turnkey clinical informatics are desired.

[0060] Dysphagia refers to different swallowing disorders and may arise secondary to stroke, multiple sclerosis, and eosinophilic esophagitis, among many other conditions. If unmanaged, dysphagia may lead to aspiration pneumonia in which food and liquid enter the airway and into lungs. The videofluoroscopic swallowing study (VFSS) is the gold standard method for dysphagia detection. In this method, clinicians detect dysphagia using a lateral X-ray video recorded during ingestion of a barium-coated bolus. The health of a swallow is judged according to criteria such as the depth of airway invasion and the degree of bolus clearance after the swallow. However, this technique requires expensive and specialized equipment, ionizing radiation and significant human resources, thereby precluding its use in the daily monitoring of dysphagia. Swallowing accelerometry has been proposed as a potential adjunct to VFSS. In this method, the patient wears a dual-axis accelerometer infero-anterior to the thyroid notch. Swallowing events are automatically extracted from the recorded acceleration signals and pattern classification methods are then deployed to discriminate between healthy and unhealthy swallows.

[0061] Some recent approaches have demonstrated the ability to automatically detect and segment distinct swallowing events, such as described in co-pending United States Patent Application Publication No. 2010/0160833, the entire contents of which are incorporated herein by reference. While this approach may provide a useful contribution

in the development of a self-standing aspiration detection device that would avoid the need for manual segmentation, most attempts at automatically classifying such swallowing events, which are generally manually segmented or distinctly recorded, have proven less fruitful. Irrespective of automatic or manual segmentation, reputation-based classification can be applied to swallowing event data signals, as discussed below.

[0062] Various features have been identified in cervical accelerometry data that can provide some discriminatory potential in classifying swallows. These include statistical features such as dispersion ratio and normality, time-frequency features such as wavelet energies, information theoretic features such as entropy rate, temporal features such as signal memory, and spectral features such as the spectral centroid. Further, in some embodiments, complementary measurement modalities, such as nasal air flow and submental mechanomyography may enhance segmentation and classification.

[0063] Given the presence of multiple feature genres and different measurement modalities, the swallow detection and classification problem lends itself to a multi-classifier approach. For example, in one embodiment, one classifier may be dedicated to each feature genre. Moreover, data sets from different patient groups may be classified using different classifiers. Furthermore, the use of multiple classifiers may be preferred in reaching ever greater classification speeds.

[0064] In this example, an exemplary embodiment of the reputation-based classifier described above was applied in automatically classifying dual-axis accelerometric signals from adult patients into safe and unsafe swallows, as labeled via videofluoroscopic review. In doing so, multiple feature genres were considered from both anterior-posterior (AP) and superior-inferior (SI) axes, and that, with respect to a relatively large data set.

[0065] In conducting this study, 30 patients were recruited (aged 65.47 \pm 13.4 years, 15 male) with suspicion of neurogenic dysphagia and who were referred to routine videofluoroscopic examination. Patients had dysphagia secondary to stroke, acquired brain injury, neurodegenerative disease, and spinal cord injury.

[0066] The data collection set-up is shown in Figure 4. Sagittal plane videofluoroscopic images of the cervical region were recorded to computer at a nominal 30 frames per second via an analog image acquisition card (PCI-1405, National Instruments). Each frame was marked with a timestamp via a software frame counter. A dual-axis accelerometer (ADXL322, Analog Devices) was taped to the participants' neck at the level of the cricoid cartilage. The axes of the accelerometer were aligned to the anatomical anterior-posterior (AP) and superior-inferior (SI) axes. Signals from both the AP and SI axes were passed through separate pre-amplifiers each with an internal bandpass filter (Model P55, Grass Technologies). The cutoff frequencies of the bandpass filter were set at 0.1 Hz and 3 kHz. The amplifier gain was 10. The signals were then sampled at 10 kHz using a data acquisition card (USB NI-6210, National Instruments) and stored on a computer for subsequent analyses. A trigger was sent from a custom LabView virtual instrument to the image acquisition card to synchronize videofluoroscopic and accelerometric recordings.

[0067] Each participant swallowed a minimum of two or a maximum of three 5mL teaspoons of thin liquid barium (40 w/v suspension) while his/her head was in a neutral position. The number of sips that the participant performed was determined by the attending clinician. The recording of dual axis accelerometry terminated after the participant finished his/her swallows. However, the participant's speech-language pathologist continued the videofluoroscopy protocol as per usual. In total, 224 individual swallowing samples were obtained from the 30 participants, 164 of which were labeled as unsafe swallows and 60 as safe swallows.

[0068] To segment the data for analysis, a speech-language pathologist reviewed the videofluoroscopy recordings. The beginning of a swallow was defined as the frame when the liquid bolus passed the point where the shadow of the mandible intersects the tongue base. The end of the swallow was identified as the frame when the hyoid bone returned to its rest position following bolus movement through the upper esophageal sphincter. The beginning and end frames as defined above were marked within the video recording using a custom C++ program. The cropped video file was then exported together with the

associated segments of dual-axis acceleration data. An unsafe swallow was defined as any swallow without airway clearance.

[0069] It has been shown that the majority of power in a swallowing vibration lies below 100Hz. Therefore, all signals were downsampled to 1KHz. Vocalization was removed from each segmented swallow according to a known periodicity detector. Whitening of the accelerometry signals to account for instrumentation nonlinearities was achieved using inverse filtering. The signals were denoised using a Daubechies-8 wavelet (8db) transform with soft thresholding. Both the decomposition level and the wavelet coefficient were chosen empirically to minimize noise while maximizing the information that remained in the signal. Figures 5 and 6 exemplify pre-processed safe and unsafe swallowing signals, respectively.

[0070] Upon completion of the above pre-processing steps, various signal features were considered for extraction, including features from multiple domains. The different genres of features are summarized below, where S is a pre-processed acceleration time series, $S = \{S1, S2, \dots, Sn\}$.

Time Domain Features

[0071] The following provides a non-exhaustive list of time domain features that may be useful, in accordance with different embodiments of the invention, in classifying cervical accelerometry data.

[0072] The sample mean is an unbiased estimation of the location of a signal's amplitude distribution and is given by

$$\mu_s = \frac{1}{n} \sum_{i=1}^n s_i.$$

[0073] The variance of a distribution measures its spread around the mean and reflects the signal's power. The unbiased estimation of variance can be obtained as

$$\sigma_s^2 = \frac{1}{n-1} \sum_{i=1}^n (s_i - \mu_s)^2.$$

[0074] The median is a robust location estimate of the amplitude distribution. For the sorted set S , the median can be calculated as

$$MED(s) = \begin{cases} s_{v+1}, & \text{if } n = 2v + 1 \\ \frac{s_v + s_{v+1}}{2}, & \text{if } n = 2v \end{cases}$$

5 [0075] Skewness is a measure of the symmetry of a distribution. This feature can be computed as follows

$$\gamma_{1,s} = \frac{\frac{1}{n} \sum_{i=1}^n (s_i - \mu_s)^3}{\left(\frac{1}{n} \sum_{i=1}^n (s_i - \mu_s)^2\right)^{1.5}}$$

[0076] A peakedness feature which reflects the peakedness of a distribution can be found as

$$\gamma_{2,s} = \frac{\frac{1}{n} \sum_{i=1}^n (s_i - \mu_s)^4}{\left(\frac{1}{n} \sum_{i=1}^n (s_i - \mu_s)^2\right)^2}$$

10

Frequency Domain Features

[0077] The following provides a non-exhaustive list of frequency domain features that may be useful, in accordance with different embodiments of the invention, in classifying cervical accelerometry data.

15 [0078] The peak magnitude value of the Fast Fourier Transform (FFT) of the signal S provides a usable frequency domain feature, wherein all the FFT coefficients are normalized by the length of the signal, n .

[0079] Another feature includes the centroid frequency of the signal S , estimated as

$$\hat{f} = \frac{\int_0^{f_{max}} f |F_s(f)|^2 df}{\int_0^{f_{max}} |F_s(f)|^2 df}$$

where $F_s(f)$ is the Fourier transform of the signal S and f_{max} is the Nyquist frequency (5KHz in this study).

[0080] Another feature includes the bandwidth of the spectrum computed using the
5 following formula

$$BW = \sqrt{\frac{\int_0^{f_{max}} (f - \hat{f})^2 |F_s(f)|^2 df}{\int_0^{f_{max}} |F_s(f)|^2 df}}$$

Information Theory-Based Features

[0081] The following provides a non-exhaustive list of information theory-based features that may be useful, in accordance with different embodiments of the invention, in
10 classifying cervical accelerometry data.

[0082] One such feature includes the entropy rate of a signal, which quantifies the extent of regularity in that signal. The measure is useful for signals with some relationship among consecutive signal points. To apply this feature, the signal S is first normalized to zero-mean and unit variance. Then, the normalized signal is quantized into
15 10 equally spaced levels, represented by the integers 0 to 9, ranging from the minimum to maximum value. Now, the sequence of U consecutive points in the quantized signal,

$$\hat{S} = \{\hat{s}_1, \hat{s}_2, \dots, \hat{s}_U\}$$

can be coded using the following equation

$$a_i = \hat{s}_{i+U-1} \cdot 10^{U-1} + \dots + \hat{s}_i \cdot 10^0$$

20 with $i = 1, 2, \dots, n-U+1$. The coded integers comprise the coding set $A_U = \{a_1, \dots, a_{n-U+1}\}$.
Using the Shannon entropy formula, entropy can be estimated as

$$E(U) = - \sum_{t=0}^{10^U-1} P_{AU}(t) \ln P_{AU}(t)$$

where $P_{AU}(t)$ represents the probability of observing the value t in A_U , approximated by the corresponding sample frequency. Then, the entropy rate can be normalized using the following equation

5
$$NE(U) = \frac{E(U) - E(U - 1) + E(1) \cdot \beta}{E(1)}$$

where β is the percentage of the coded integers in A_L that occurred only once. Finally, the regularity index $\rho \in [0,1]$ can be obtained as

$$\rho = 1 - \min NE(U)$$

10 where a value of ρ close to 0 signifies maximum randomness while ρ close to 1 indicates maximum regularity.

[0083] Another feature is the signal's memory. To calculate the memory of the signal, its autocorrelation function can be computed from zero to the maximum time lag and normalized such that the autocorrelation at zero lag is unity. The memory can be estimated as the time required for the autocorrelation to decay to $1/e$ of its zero lag value.

15 [0084] Another feature is the Lempel-Ziv (L-Z) complexity, which measures the predictability of a signal. To compute the L-Z complexity for signal S , first, the minimum and the maximum values of signal points can be calculated and then, the signal can be quantized into 100 equally spaced levels between its minimum and maximum values. Then, the quantized signal,

20
$$B_1^n = \{b_1, b_2, \dots, b_n\}.$$

can be decomposed into T different blocks,

$$B_1^n = \{\psi_1, \psi_2, \dots, \psi_T\}$$

[0085] A block Ψ can be defined as

$$\Psi = B_j^\ell = \{b_j, b_{j+1}, \dots, b_\ell\}, 1 \leq j \leq \ell \leq n$$

and values thereof can be calculated as follows:

$$\Psi = \begin{cases} \psi_m = b_1, & \text{if } m=1, \\ \psi_{m+1} = B_{h_m+1}^{h_{m+1}}, & m \geq 1, \end{cases}$$

5 where h_m is the ending index for ψ_m , such that ψ_{m+1} is a unique sequence of minimal length within the sequence

$$B_1^{h_{m+1}-1}$$

Finally, the normalized L-Z complexity can be calculated as

$$LZ = \frac{T \log_{100} n}{n}$$

10 [0086] As will be appreciated by the skilled artisan, different subsets and combinations of the above features, as well as other features, can be used in different embodiments to classify accelerometry data, without departing from the general scope and nature of the present disclosure.

[0087] The signal features introduced above were ranked using the Fisher ratio for
 15 univariate separability. In the time domain, mean and variance in the AP axis and skewness in the SI axis were the top-ranked features. Similarly, in the frequency domain, the peak magnitude of the FFT and the spectral centroid in the AP direction and the bandwidth in the SI direction were retained. Finally, in the information theoretic domain, entropy rate for the SI signal and memory of the AP signal were the highest ranking
 20 features. Consideration was subsequently limited to these feature subsets for classification. For comparison between single and dual-axes classifiers, classifiers that employed feature subsets (as identified above) from a single axis were also considered.

[0088] Given the disproportion of safe and unsafe samples, a smooth bootstrapping procedure was invoked to balance the classes. All features were then standardized to 0 mean and unit variance. Three separate support vector machine (SVM) classifiers were invoked, one for each feature genre (time, frequency and information theoretic). Hence, 5 the feature space dimensionalities for the classifiers were 3 (SVM with time features), 3 (SVM with frequency features) and 2 (SVM with information-theoretic features). The use of different feature sets for each classifier generally ensures that the classifiers will perform independently.

[0089] Classifier accuracy was estimated via a 10-fold cross validation with a 90-10 10 split. In each fold, the whole training set was used to estimate the individual classifier reputations. Classifiers were then ranked according to their reputation values. Without loss of generality, assume $r_1 \geq r_2 \geq r_3$. If θ_1 and θ_2 cast the same vote about a test swallow, their common decision was accepted as the final classification. However, if they voted differently, the a posteriori probability of each class was computed and the 15 maximum a posteriori probability rule was applied to select the final classification.

[0090] The sensitivity, specificity and accuracy of the single-axis and dual-axis accelerometry classifiers are summarized in Figure 7. The dual-axis classifier had significantly higher accuracy (80.48+/-5.0) than either single-axis classifier ($p < 0.05$, two-sample t-test), specificity (64+/-8.8) comparable to that of the SI classifier ($p=1.0$) 20 and sensitivity (97.1+/-2) on par with that of the AP classifier ($p=1.0$). In other words, the dual-axis classifier retained the best sensitivity and specificity achievable with either single-axis classifier.

[0091] Of the two axes, the AP axis tended to carry more useful information than the SI direction for discrimination between safe and unsafe swallowing. This observation is 25 evidenced in Figure 7, where AP accuracy is higher than SI levels. Nonetheless, the SI axis does carry information distinct from that of the AP orientation, as dual-axis classification exceeds any single-axis counterpart. Results thus support the inclusion of selected features from both the AP and SI axes for the automatic discrimination between safe and unsafe swallowing.

[0092] In a recent videofluoroscopic study, both AP and SI accelerations were attributed to the planar motion of the hyoid and larynx during swallowing. In that study, the displacement of the hyoid bone and larynx along with their interaction explained over 70% of the variance in the doubly integrated acceleration in both AP and SI axes at the level of the cricoid cartilage. Juxtaposed with the above findings above, this reported physiological source of swallow accelerometry suggests that it is the difference in hyolaryngeal motion that is manifested as discriminatory cues between safe and unsafe swallowing. Indeed, early single-axis accelerometry research had implicated decreased laryngeal elevation as the reason for suppressed AP accelerations in individuals with severe dysphagia.

[0093] Figure 8 is a parallel axes plot depicting the internal representation of safe and unsafe swallows acquired by the reputation-based classifier. Each feature has been normalized by its standard deviation to facilitate visualization. On each axis, the range of values between the first and third quartile of the feature values are shown with a horizontal line. The quartile values of adjacent axes are joined by solid (safe swallow) or dashed (unsafe swallow) lines. From this, distinct patterns are observed which characterize each type of swallow. Unsafe swallows tend to have lower mean acceleration amplitude, narrower variance, higher spectral centroid and longer memory. The lower mean vibration amplitude in unsafe swallowing resonates with previous reports of suppressed peak acceleration in dysphagic patients and reduced peak anterior hyoid excursion in older adults, both suggesting compromised airway protection. Similarly, the narrower variance implies a contracted dynamic range of hyolaryngeal acceleration in unsafe swallowing. The observation of a higher spectral centroid in unsafe swallowing may reflect departures from the typical axial high-low frequency coupling trends of normal swallowing. The longer memory and hence slower decay of the autocorrelation may be indicative of inherent non-stationarities in unsafe swallowing.

[0094] Unsafe swallows are also noted to be negatively skewed while safe swallows are evenly split between positive and negative skew. In other words, in unsafe swallowing, the upward motion of the hyolaryngeal structure appears to have weaker accelerations than during the downward motion. This is opposite of the previously

reported tendency for healthy swallowing and may reflect inadequate urgency to protect the airway.

[0095] The merit of a reputation-based classifier for the present problem can be appreciated by contrasting its performance against that of the classic method of combining classifiers, i.e., via the majority voting algorithm. To this end, Figure 9 summarizes the accuracies of both approaches from a 10 fold cross-validation using the data of this study. Clearly, the location of the density of reputation-based accuracies appears to be further to the right of the location of the majority voting density. The large spread in both densities amplifies the risk of Type II error and thus conventional testing (e.g., Wilcoxon ranksum) fails to identify any differences. However, upon more careful inspection using a two-sample Kolmogorov-Smirnoff test of the 20% one-sided trimmed densities (i.e., omitting the 2 most extreme points in each density), a statistically significant difference between the distributions is confirmed.

[0096] This study has demonstrated the potential for automatic discrimination between safe and unsafe (without airway clearance) swallows on the basis of a selected subset of time, frequency and information theoretic features derived from non-invasive, dual-axis accelerometric measurements at the level of the cricoid cartilage. Dual-axis classification was more accurate than single-axis classification. The reputation-based classifier internally represented unsafe swallows as those with lower mean acceleration, lower range of acceleration, higher spectral centroid, slower autocorrelation decay and weaker acceleration in the superior direction. Reputation-based classification of dual-axis swallowing accelerometry was shown to present an advantageous solution over previous classification techniques in implementing a turn-key clinical assessment device.

EXAMPLE 2

[0097] The above-described classification methods are applied, as above, and in accordance with another exemplary embodiment of the invention, to the classification of healthy and unhealthy swallows. Specifically, this example is set to differentiate between safe and unsafe swallowing on the basis of dual-axis accelerometry. The basic idea is to decompose a high dimensional classification problem into 3 lower dimensional problems,

each with a unique subset of features and a dedicated classifier. The individual classifier decisions are then melded according to the described reputation algorithm.

[0098] In this example, a randomly selected subset of 100 healthy swallows and 100 dysphagic swallows were selected from an existing database of accelerometric data, with similar pre-processing approaches applied thereto as discussed above.

[0099] In this example, 3 separate back-propagation neural network (NN) classifiers were trained, one for each genre of signal feature outlined above. Hence, the feature space dimensionalities for the classifiers were 4 (NN with time features), 3 (NN with frequency features) and 3 (NN with information-theoretic features). Each neural network classifier had 2 inputs, 4 hidden units and 1 output. Although it is possible to invoke different classifiers for each genre of signal feature, the same classifiers were utilized in this example to facilitate the evaluation of local decisions. The use of different feature sets for each classifier generally ensures that the classifiers will perform independently.

[00100] Consistent with above description, first, the three small neural networks, classify their inputs independently. Then, using the outputs of these classifiers and their respective reputation values, the reputation-based method determines the correct label of the input. Classifier accuracy was estimated via a 10-fold cross validation with a 90-10 split. However, unlike classical cross-validation, the 'training' set was further segmented into an actual training set and a validation set. In other words, in each fold, 160 (80%) swallows were used for training, 20 (10%) for validation and 20 (10%) reserved for testing. Among the 20 swallows of the validation set, 10 were used as a traditional validation set and 10 were used for computation of the reputation values. After training, classifier reputations were estimated using this second validation set. Classifiers were then ranked according to their reputation values.

[00101] As in the above example, and without loss of generality, assume $r_1 \geq r_2 \geq r_3$. If θ_1 and θ_2 cast the same vote about a test swallow, their common decision was accepted as the final classification. However, if they voted differently, the a posteriori probability of each class was computed and the maximum a posteriori probability rule was applied to select the final classification. To better understand the difference between the multiple

classifier system and a single, all-encompassing classifier, a multilayer neural network was also trained via back-propagation with all 10 features, i.e., using the collective inputs of all three smaller classifiers. This all-encompassing classifier, from hereon referred to as the grand classifier, also had 4 hidden units. The accuracies of the individual classifiers were also statistically compared against those of a majority vote classifier combination and a reputation-based classifier combination.

[00102] Table 1 tabulates the local and global classification results. On average, the frequency domain classifier appears best among the individual NNs while the information-theoretic NN fairs worst. Also, it is clear from this table that by combining the local decisions of the classifiers, using a reputation-based method as described above, the overall performance of the system increases. The result of the grand classifier is statistically the same as the small classifiers. However, training this classifier is more difficult and requires more time, thus making this approach of little value. Collectively, these results indicate that there is merit in combining neural network classifiers in this problem domain. The accuracy of the majority vote neural network combination did not significantly differ from that of the individual ($p > 0.11$) and grand classifiers ($p=0.16$). On the other hand, the reputation-based combination led to further improvement in accuracy over the time domain ($p=0.04$) and information-theoretic ($p=0.05$) classifiers, but did not significantly surpass the grand ($p=0.09$) and frequency domain networks ($p=0.09$). The reputation-based scheme yields accuracies better than those previously reported using alternate methods (74%), wherein the entire database was required and the maximum feature space dimension was 12. In this example, only a fraction of the database was considered and no classifier had a feature space dimensionality greater than 4. Therefore, the system considered in this example offers the advantages of computational efficiency and less stringent demands on training data. Accordingly, the merits of applying a reputation-based neural network combination for classification of a dysphagia dataset is confirmed.

Table 1. The average performance of the individual classifiers and their reputation-based combination.

Classifier	Average Performance (%)
Time domain NN	67.5 ± 12.5
Frequency domain NN	69.5 ± 10.3
Information theoretic NN	65.5 ± 11.2
Grand classifier	70.0 ± 8.5
Majority vote	74.5 ± 8.9
Combined classifier decision	78.0 ± 8.2

[00103] While the present disclosure describes various exemplary embodiments, the disclosure is not so limited. To the contrary, the disclosure is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims. The scope of the following claims is to be accorded the broadest interpretation so as to encompass all such modifications and equivalent structures and functions.

CLAIMS:

1. A method for classifying test data using two or more classifiers, the method comprising the steps of:
 - training said two or more classifiers using a training data set;
 - 5 measuring a respective overall performance of each of said trained classifiers using a validation data set;
 - assigning a respective reputation value to each of said trained classifiers representative of said respective overall performance thereof; and
 - 10 classifying the test data via combination of said two or more trained classifiers as a function of said respective reputation values.
2. The method of claim 1, wherein said overall performance measure consists of an overall accuracy of each said trained classifier in classifying said validation data set.
- 15 3. The method of claim 1, wherein said overall performance measure consists of a percentage of correct classifications by each said trained classifier on said validation set.
4. The method of any one of claims 1 to 3, wherein each said reputation value is assigned as a function of measured abstract level classifier outputs over said validation
20 set.
5. The method of claim 4, wherein each said trained classifier comprises an abstract level classifier.
- 25 6. The method of anyone of claims 1 to 5, wherein each said reputation value is independently calculated from one another.
7. The method of any one of claims 1 to 6, said classifying step comprising:
 - for each given class, calculating a likelihood that said given class is correct as a
30 function of each classifier output and each said reputation value; and

selecting a highest likelihood output class as a global classification for the test data.

8. The method of claim 7, wherein said likelihood is calculated as a function of said reputation value where a given classifier output coincides with said given class, and as a function of a complement of said reputation value otherwise.

9. The method of claim 8, wherein said likelihood is calculated in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t)}$$

10 where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the likelihood that class ω_j is a correct global classification given classifier outputs θ_1 to θ_L , and where $p(\theta_i | \omega_j)$ is equal to a reputation value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this reputation value otherwise.

10. The method of claim 8, wherein said likelihood is calculated in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j) p(\omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t) p(\omega_t)}$$

where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the likelihood that class ω_j is a correct global classification given classifier outputs θ_1 to θ_L , where $p(\theta_i | \omega_j)$ is equal to a reputation value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this reputation value otherwise, and where $p(\omega_t)$ is a prior probability of class ω_t .

11. The method of any one of claims 7 to 10, wherein said classifying step further comprises, prior to said likelihood calculating step, classifying the test data using each of said two or more classifiers to obtain respective classifier outputs;

comparing said respective classifier outputs for a subset of said classifiers having highest respective reputation values; and

upon each of said respective classifier outputs in said subset coinciding with a same output, outputting said same output as the global classification for the test data;

5 otherwise

proceeding with said likelihood calculating step.

12. The method of any one of claims 1 to 11, wherein said test data comprises cervical accelerometry data, and wherein said classifiers are trained to classify said
10 cervical accelerometry data as representative of one of a healthy and an unhealthy swallowing event.

13. The method of claim 12, wherein said classifiers comprise at least two of a time domain classifier, a frequency domain classifier and an information theory domain
15 classifier.

14. The method of claim 13, wherein said time-domain classifier is trained to classify at least one time-domain feature selected from a mean, a variance, a median, a skewness and a peakedness of the test data.
20

15. The method of claim 13 or 14, wherein said frequency-domain classifier is trained to classify at least one frequency-domain feature selected from a peak magnitude, a centroid frequency and a bandwidth of the test data.

25 16. The method of any one of claims 13 to 15, wherein said information theory domain classifier is trained to classify at least one information theory domain feature selected from an entropy, a memory and a Lempel-Ziv complexity of the test data.

30 17. The method of any one of claims 13 to 16, wherein said cervical accelerometry data comprises dual-axis cervical accelerometry data.

18. The method of claim 17, wherein different classifiers are selected for locally classifying data acquired via different axes.
19. The method of any one of claims 7 to 18, wherein said classifying step is
5 automatically implemented by a computing device comprising a processor and a computer-readable medium associated therewith, said computer-readable medium having stored thereon statements and instructions to be implemented by said processor in implementing said classifying step.
- 10 20. The method of any one of claims 1 to 18, automatically implemented by a computing device configured to receive as input said training data set, validation data set and test data, and comprising a processor and a computer-readable medium associated therewith, said computer-readable medium having stored thereon statements and instructions to be implemented by said processor in implementing the method to output a
15 classification of the test data.
21. The method of any one of claims 1 to 20, wherein said training data set is distinct from said validation data set.
- 20 22. A method for classifying test data using two or more classifiers, the method comprising the steps of:
classifying the test data using each of said two or more classifiers to obtain respective classifications therefrom;
calculating a highest likelihood classification for the test data as a function of said
25 respective classifications and as a function of a respective overall performance value previously measured for each of the two or more classifiers; and
outputting said highest likelihood classification as global classification for the test data.
- 30 23. The method of claim 22, wherein each said overall performance value consists of an overall accuracy of a given classifier in classifying a known data set.

24. The method of claim 22, wherein each said overall performance value consists of a percentage of correct classifications by a given classifier in classifying a known data set.

5

25. The method of any one of claims 22 to 24, wherein each of the classifiers comprises an abstract level classifier.

26. The method of any one of claims 22 to 25, said classifying step comprising:

10 for each given class, calculating a respective likelihood that said given class is correct as a function of said respective classifications and each said respective overall performance value; and

selecting said highest likelihood classification therefrom for output as said global classification.

15

27. The method of claim 26, wherein each said respective likelihood is calculated as a function of said respective overall performance value where a given classifier output coincides with said given class, and as a function of a complement of said respective overall performance value otherwise.

20

28. The method of claim 26, wherein each said respective likelihood is calculated in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t)}$$

where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the respective likelihood that class ω_j is a correct global classification given classifier outputs θ_1 to θ_L , and where $p(\theta_i | \omega_j)$ is equal to said

25 respective overall performance value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this performance value otherwise.

29. The method of claim 25, wherein said likelihood is calculated in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j) p(\omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t) p(\omega_t)}$$

where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the respective likelihood that class ω_j is a correct global
 5 classification given classifier outputs θ_1 to θ_L , where $p(\theta_i | \omega_j)$ is equal to said respective overall performance value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this performance value otherwise, and where $p(\omega_t)$ is a prior probability of class ω_t .

10 30. The method of any one of claims 21 to 29, further comprising prior to said calculating step, the steps of:

comparing said respective classifications for a subset of the two or more classifiers having highest respective overall performance values; and

upon each of said respective classifications in said subset coinciding with a same
 15 output, selecting said same output for output as said global classification; otherwise proceeding with said calculating step.

31. The method of any one of claims 21 to 30, wherein said test data comprises cervical accelerometry data, and wherein said classifiers are trained to classify said
 20 cervical accelerometry data as representative of one of a healthy and an unhealthy swallowing event.

32. The method of claim 31, wherein said classifiers comprise at least two of a time domain classifier, a frequency domain classifier and an information theory domain
 25 classifier.

33. The method of claim 32, wherein said time-domain classifier is trained to classify at least one time-domain feature selected from a mean, a variance, a median, a skewness and a peakedness of the test data.
- 5 34. The method of claim 32 or 33, wherein said frequency-domain classifier is trained to classify at least one frequency-domain feature selected from a peak magnitude, a centroid frequency and a bandwidth of the test data.
35. The method of any one of claims 32 to 34, wherein said information theory
10 domain classifier is trained to classify at least one information theory domain feature selected from an entropy, a memory and a Lempel-Ziv complexity of the test data.
36. The method of any one of claims 32 to 35, wherein said cervical accelerometry data comprises dual-axis cervical accelerometry data.
- 15 37. The method of claim 36, wherein different classifiers are selected for locally classifying data acquired via different axes.
38. The method of any one of claims 22 to 37, automatically implemented by a
20 computing device configured to receive as input said test data, and comprising a processor and a computer-readable medium associated therewith, said computer-readable medium having stored thereon statements and instructions to be implemented by said processor in implementing the method to output said global classification.
- 25 39. A computer-readable medium having statements and instructions stored thereon that, upon execution by a processor, automatically implement the steps of any one of claims 1 to 38.
40. A computer-readable medium having statements and instructions stored thereon
30 for execution by a processor of a computing device in automatically classifying input test data, the statements and instructions comprising:

two or more encoded classifiers each configured to output respective local data classifications;

a training module for training said two or more classifiers on a training data set;

5 a validation module for measuring a respective overall performance value for each of said trained classifiers using a validation data set, and assigning a respective reputation value to each of said trained classifiers as a function thereof; and

10 a classification module for locally classifying the test data via each of said two or more trained classifiers, and globally classifying the test data as a function of each said respective reputation value and said respective local data classifications output from said trained classifiers on the test data.

41. The computer-readable medium of claim 40, wherein said validation module is configured to compute an overall accuracy of each of said trained classifiers in classifying said validation data set, and to define each said respective reputation value as
15 a function thereof.

42. The computer-readable medium of claim 40, wherein said validation module is configured to compute a percentage of correct classifications by each of said trained classifiers in classifying said validation data set, and to define each said respective
20 reputation value as a function thereof.

43. The computer-readable medium of any one of claims 40 to 42, said encoded classifiers comprising abstract level classifiers.

25 44. The computer-readable medium of any one of claims 40 to 43, wherein said classification module comprises statements and instructions for:

calculating, for each given class, a likelihood that said given class is correct as a function of each of said trained classifier outputs on the test data and each said respective reputation value; and

30 selecting a highest likelihood output class as a global classification for the test data.

45. The computer-readable medium of claim 44, wherein said likelihood is calculated as a function of said respective reputation value where a given trained classifier output coincides with said given class, and as a function of a complement of said respective
5 reputation value otherwise.

46. The computer-readable medium of claim 45, wherein said likelihood is calculated in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t)}$$

where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the likelihood that class ω_j is a correct global classification
10 given classifier outputs θ_1 to θ_L , and where $p(\theta_i | \omega_j)$ is equal to a reputation value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this reputation value otherwise.

47. The computer readable medium of claim 45, wherein said likelihood is calculated
15 in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j) p(\omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t) p(\omega_t)}$$

where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the likelihood that class ω_j is a correct global classification
20 given classifier outputs θ_1 to θ_L , where $p(\theta_i | \omega_j)$ is equal to a reputation value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this reputation value otherwise, and where $p(\omega_t)$ is a prior probability of class ω_t .

48. The computer-readable medium of any one of claims 44 to 47, wherein said classification module further comprises statements and instructions for, prior to said likelihood calculating step;

comparing said respective trained classifier outputs on the test data for a subset of said classifiers having highest respective reputation values; and

upon each of said respective trained classifier outputs in said subset coinciding with a same output, outputting said same output as the global classification for the test data; otherwise

proceeding with likelihood calculating step.

49. The computer-readable medium of any one of claims 40 to 48, wherein said test data comprises cervical accelerometry data, and wherein said classifiers are trained to classify said cervical accelerometry data as representative of one of a healthy and an unhealthy swallowing event.

50. The computer-readable medium of any one of claims 40 to 49, wherein said training data set is distinct from said validation data set.

51. A computer-readable medium having statements and instructions stored thereon for execution by a processor of a computing device in automatically classifying input test data, the statements and instructions comprising:

two or more encoded classifiers each trained to output respective local data classifications;

a respective reputation value assigned to each of said two or more classifiers representative of a respective overall performance thereof; and

a classification module for locally classifying the test data via each of said two or more trained classifiers, and globally classifying the test data as a function of each said respective reputation value and said respective local data classifications output from said trained classifiers on the test data.

52. The computer-readable medium of claim 51, wherein each said respective reputation value consists of a previously measured overall accuracy in classifying a known data set.

53. The computer-readable medium of claim 51, wherein each said respective reputation value consists of a previously measured percentage of correct classifications in classifying a known data set.

5 54. The computer-readable medium of any one of claims 51 to 53, wherein said classification module comprises statements and instructions for:
calculating, for each given class, a likelihood that said given class is correct as a function of each said respective reputation value and said respective local data classifications output from said trained classifiers on the test data; and
10 selecting a highest likelihood output class as a global classification for the test data.

55. The computer-readable medium of claim 54, wherein a given likelihood is calculated as a function of said respective reputation value where a given trained
15 classifier output on the test data coincides with said given class, and as a function of a complement of said respective reputation value otherwise.

56. The computer-readable medium of claim 54, wherein said likelihood is calculated in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t)}$$

20 where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the likelihood that class ω_j is a correct global classification given classifier outputs θ_1 to θ_L , and where $p(\theta_i | \omega_j)$ is equal to a reputation value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this reputation value otherwise.

25 57. The computer readable medium of claim 54, wherein said likelihood is calculated in accordance with the following:

$$p(\omega_j | \theta_1, \dots, \theta_L) = \frac{\prod_{i=1}^L p(\theta_i | \omega_j) p(\omega_j)}{\sum_{t=1}^c \prod_{i=1}^L p(\theta_i | \omega_t) p(\omega_t)}$$

where $p(\omega_j | \theta_1, \dots, \theta_L)$ is the likelihood that class ω_j is a correct global classification given classifier outputs θ_1 to θ_L , where $p(\theta_i | \omega_j)$ is equal to a reputation value assigned to classifier θ_i when an output thereof coincides with ω_j and is equal to a complement of this
5 reputation value otherwise, and where $p(\omega_i)$ is a prior probability of class ω_i .

58. The computer-readable medium of any one of claims 54 to 57, wherein said classification module further comprises statements and instructions for, prior to said likelihood calculating step;
10 comparing said respective trained classifier outputs on the test data for a subset of said classifiers having highest respective reputation values; and
upon each of said respective trained classifier outputs in said subset coinciding with a same output, outputting said same output as the global classification for the test data; otherwise
15 proceeding with likelihood calculating step.

59. The computer-readable medium of any one of claims 51 to 58, wherein said test data comprises cervical accelerometry data, and wherein said classifiers are trained to classify said cervical accelerometry data as representative of one of a healthy and an
20 unhealthy swallowing event.

60. A device for classifying test data using two or more classifiers, the device comprising:
a processor;
25 an input for receiving test data to be classified;
an output for outputting a global classification of the test data;
a computer-readable data storage device operatively coupled to said processor, input and output, and having stored thereon statements and instructions for execution by said processor in classifying the test data, said statements and instructions comprising:
30 two or more encoded classifiers each trained to output respective local data classifications;

a respective reputation value assigned to each of said two or more classifiers representative of a respective overall performance thereof; and
a classification module for locally classifying the test data via each of said two or more trained classifiers, globally classifying the test data as a function of each said respective reputation value and said respective local data classifications output from said trained classifiers on the test data, and communicating a resulting global classification to said output.

5
10
61. The device of claim 60, said input further for receiving a known data set, said statements and instructions further comprising a validation module for measuring a respective overall performance value for each of said trained classifiers in classifying said known data set, and assigning said respective reputation value to each of said trained classifiers as a function thereof.

15
62. The device of claim 61, further comprising a training module for training said encoded classifiers.

20
63. The device of any one of claims 60 to 62, wherein the test data comprises cervical accelerometry data, and wherein said classifiers are trained to classify said cervical accelerometry data as representative of one of a healthy and an unhealthy swallowing event.

25
64. The device of claim 63, wherein the cervical accelerometry data comprises dual-axis cervical accelerometry data.

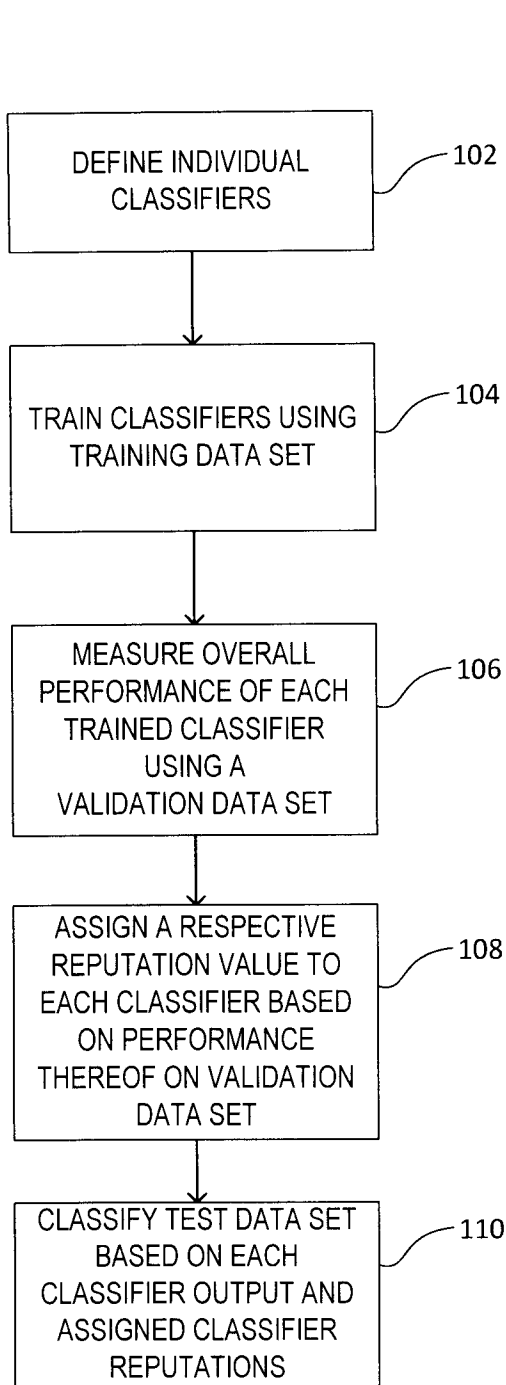


FIGURE 1

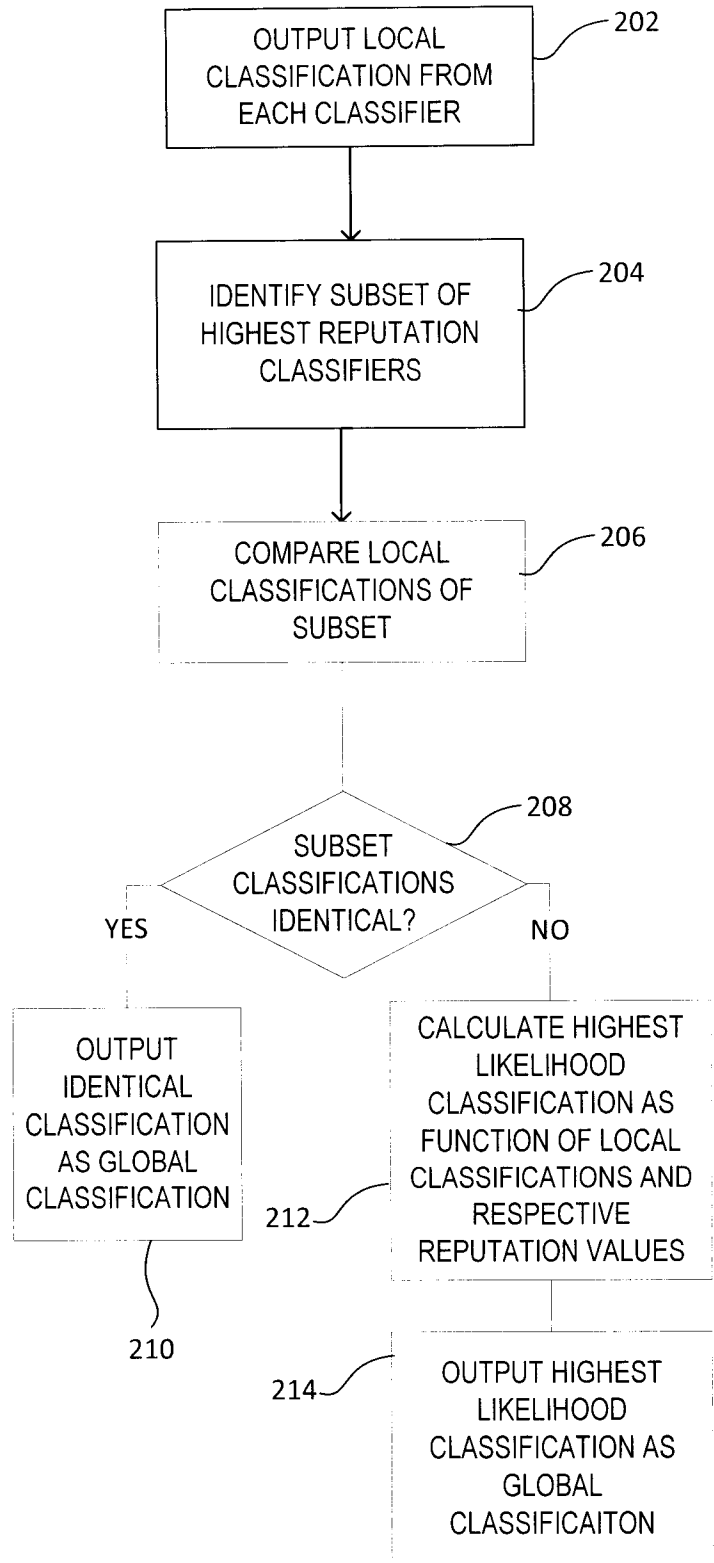


FIGURE 2

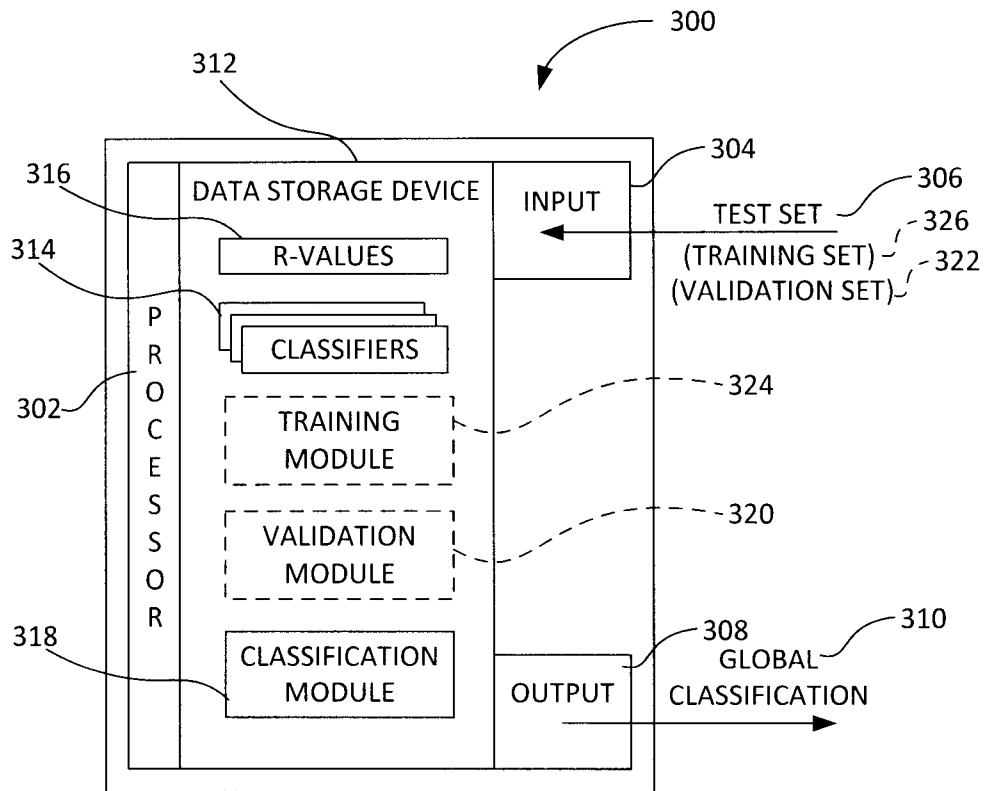


FIGURE 3

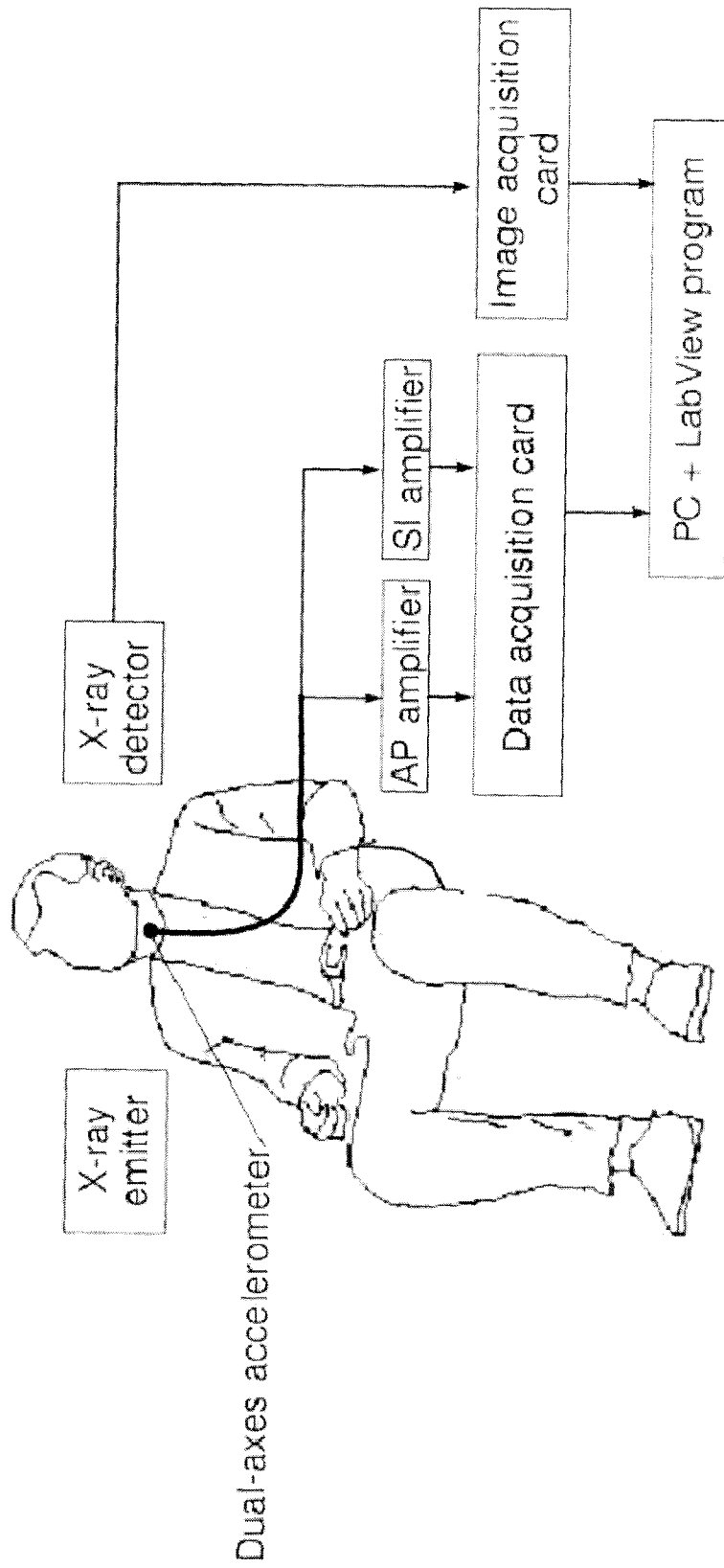


FIGURE 4

FIGURE 5

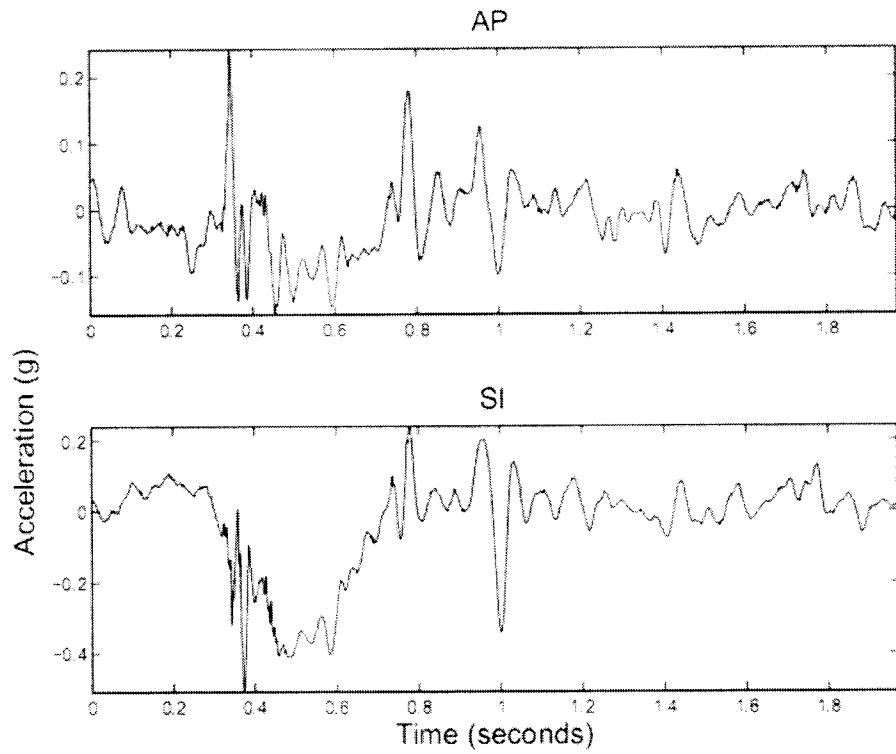
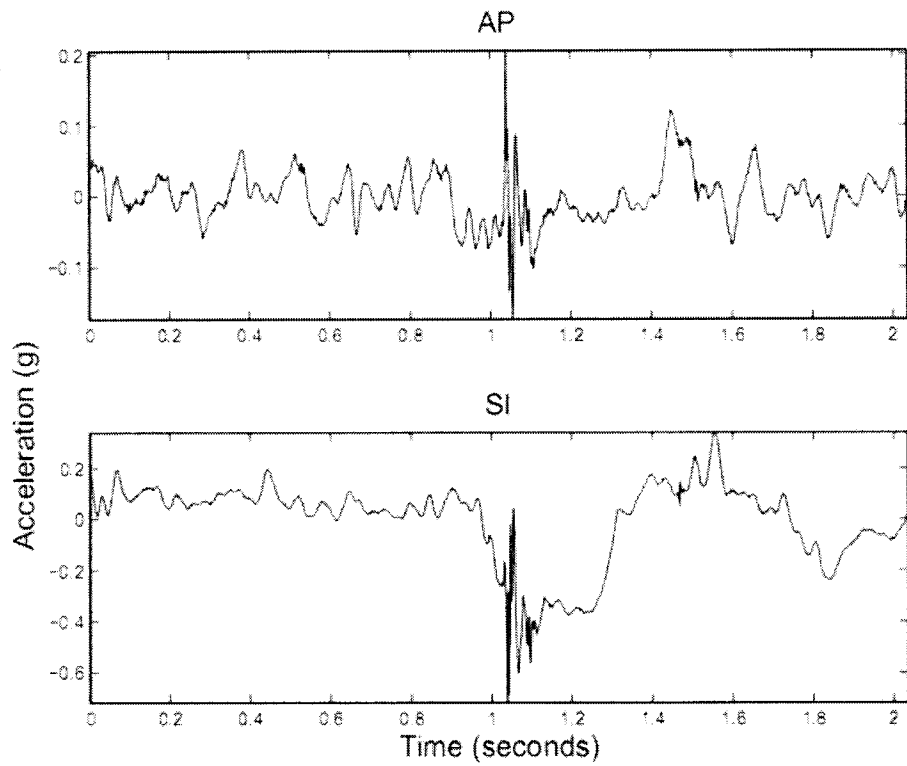


FIGURE 6



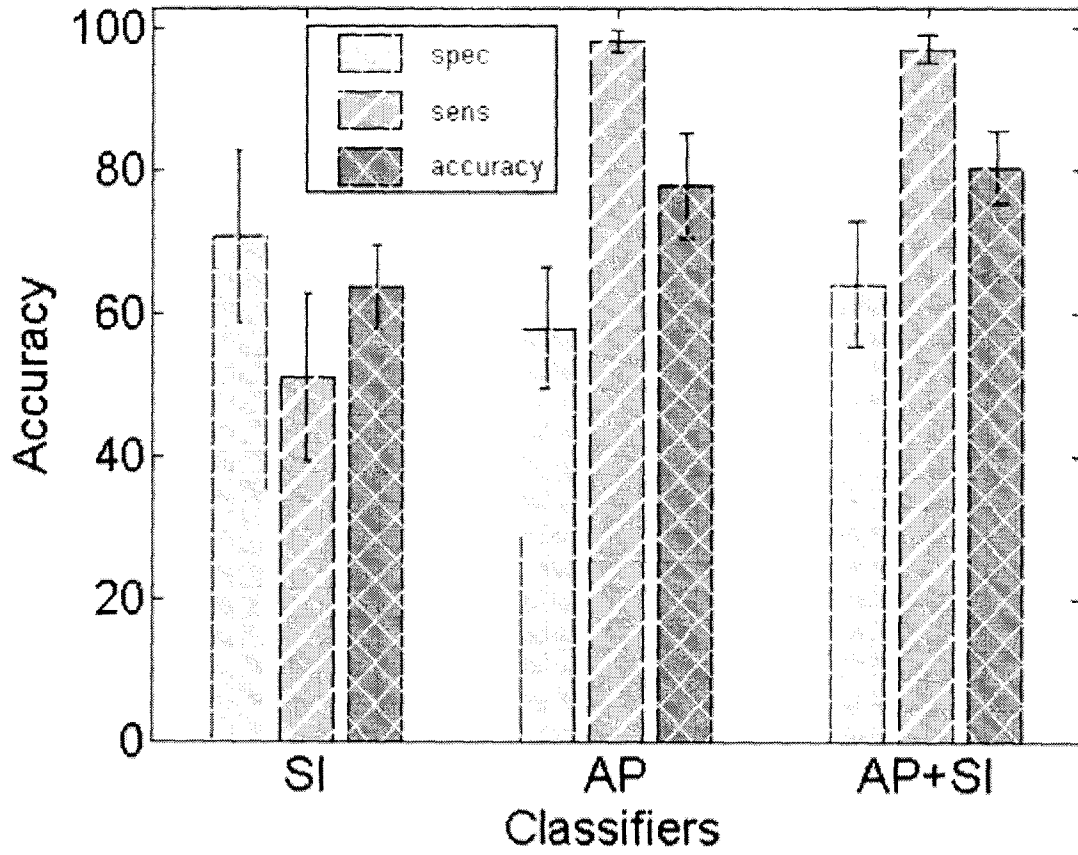


FIGURE 7

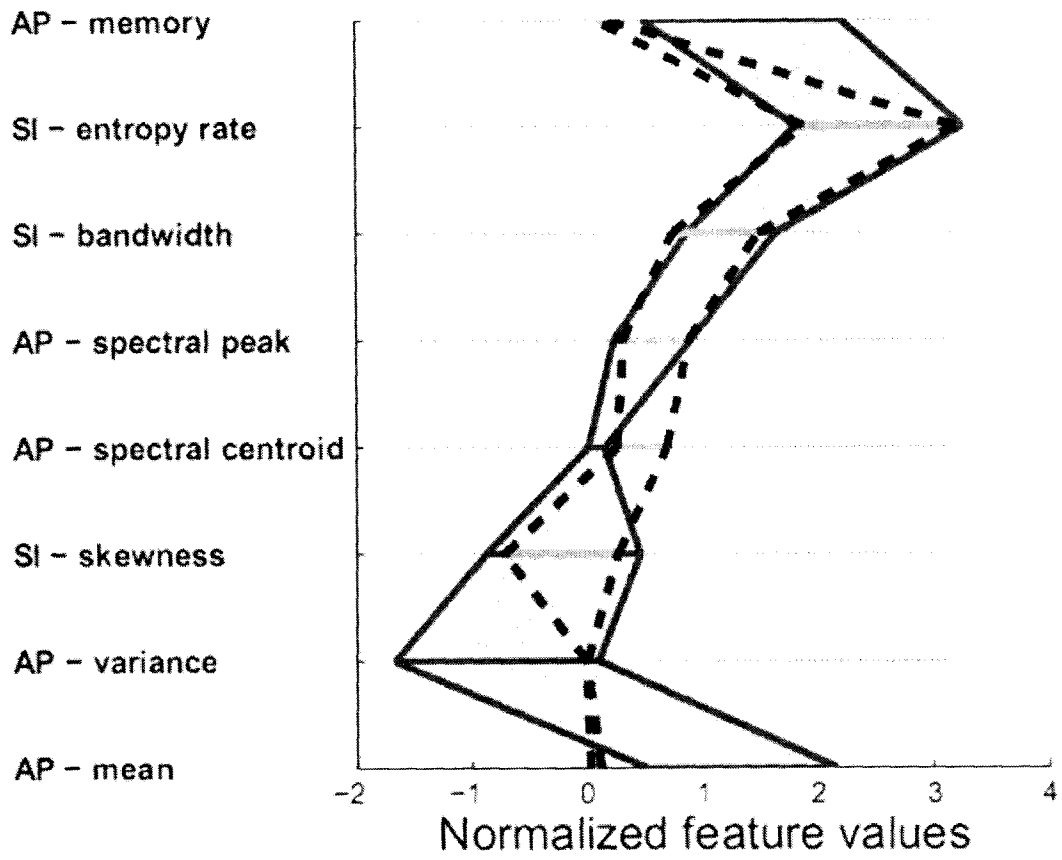


FIGURE 8

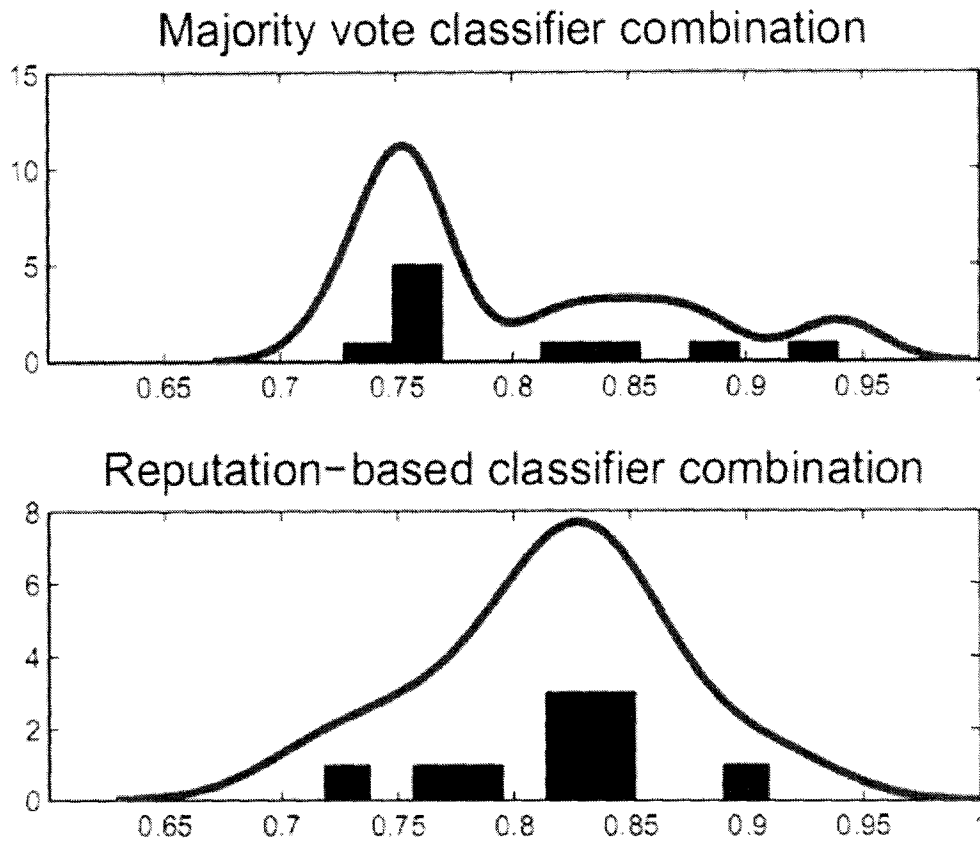


FIGURE 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2011/001085

<p>A. CLASSIFICATION OF SUBJECT MATTER IPC: G06F 17/00 (2006.01) , G06F 15/18 (2006.01) , G06F 17/10 (2006.01) According to International Patent Classification (IPC) or to both national classification and IPC</p>																										
<p>B. FIELDS SEARCHED</p> <p>Minimum documentation searched (classification system followed by classification symbols) IPC (2006): G06F (in combination with keywords)</p> <p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> <p>Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used) Database: TotalPatent, Canadian Patent Database (Intellect), EPOQUE (EPODOC), USPTO West, IEEE Xplore Keywords: classifier, training, data set, validation, test, combine, accelerometry, swallow</p>																										
<p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p> <table border="1" style="width:100%; border-collapse: collapse;"> <thead> <tr> <th style="width:10%;">Category*</th> <th style="width:60%;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="width:30%;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td align="center">X</td> <td>US 2006/0120609 (IVANOV et al.) 08 June 2006 (08-06-2006) * Paragraphs [0007], [0031], [0034]-[00431] and [0061]-[0063] *</td> <td>1-8, 11, 21-27, 30, 40-45, 48, 50-55, 58 and 60-62</td> </tr> <tr> <td align="center">---</td> <td></td> <td align="center">---</td> </tr> <tr> <td align="center">Y</td> <td></td> <td>9, 10, 12-15, 17-20, 28, 29, 31-34, 36-39, 46, 47, 49, 56, 57, 59, 63 and 64</td> </tr> <tr> <td align="center">Y</td> <td>TAX, D. et al., "Combining Multiple Classifier by a Averaging or by Multiplying?", Pattern Recognition 33, pages 1475-1485, 2000. * Page 1477 *</td> <td>9, 19, 20, 28, 38, 39, 46 and 56</td> </tr> <tr> <td align="center">Y</td> <td>KITTLER, J. et al., "On Combining Classifier", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, No. 3, March 1998. * Page 128 *</td> <td>10, 19, 20, 29, 38, 39, 47 and 57</td> </tr> <tr> <td align="center">Y</td> <td>US 2010/0160833 (CHAU et al.) 24 June 2010 (24-06-2010) * Paragraphs [0010], [0030], [0039]-[0061], [0076] and [0077] *</td> <td>12-14, 17-20, 31-33, 36-39 49, 59, 63 and 64</td> </tr> <tr> <td align="center">Y</td> <td>US 2008/0269646 (CHAU et al.) 30 October 2008 * Paragraphs [0054]-[0056] and [0090] *</td> <td>14, 15, 19, 20, 33, 34, 38 and 39</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	US 2006/0120609 (IVANOV et al.) 08 June 2006 (08-06-2006) * Paragraphs [0007], [0031], [0034]-[00431] and [0061]-[0063] *	1-8, 11, 21-27, 30, 40-45, 48, 50-55, 58 and 60-62	---		---	Y		9, 10, 12-15, 17-20, 28, 29, 31-34, 36-39, 46, 47, 49, 56, 57, 59, 63 and 64	Y	TAX, D. et al., "Combining Multiple Classifier by a Averaging or by Multiplying?", Pattern Recognition 33, pages 1475-1485, 2000. * Page 1477 *	9, 19, 20, 28, 38, 39, 46 and 56	Y	KITTLER, J. et al., "On Combining Classifier", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, No. 3, March 1998. * Page 128 *	10, 19, 20, 29, 38, 39, 47 and 57	Y	US 2010/0160833 (CHAU et al.) 24 June 2010 (24-06-2010) * Paragraphs [0010], [0030], [0039]-[0061], [0076] and [0077] *	12-14, 17-20, 31-33, 36-39 49, 59, 63 and 64	Y	US 2008/0269646 (CHAU et al.) 30 October 2008 * Paragraphs [0054]-[0056] and [0090] *	14, 15, 19, 20, 33, 34, 38 and 39
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.																								
X	US 2006/0120609 (IVANOV et al.) 08 June 2006 (08-06-2006) * Paragraphs [0007], [0031], [0034]-[00431] and [0061]-[0063] *	1-8, 11, 21-27, 30, 40-45, 48, 50-55, 58 and 60-62																								
---		---																								
Y		9, 10, 12-15, 17-20, 28, 29, 31-34, 36-39, 46, 47, 49, 56, 57, 59, 63 and 64																								
Y	TAX, D. et al., "Combining Multiple Classifier by a Averaging or by Multiplying?", Pattern Recognition 33, pages 1475-1485, 2000. * Page 1477 *	9, 19, 20, 28, 38, 39, 46 and 56																								
Y	KITTLER, J. et al., "On Combining Classifier", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, No. 3, March 1998. * Page 128 *	10, 19, 20, 29, 38, 39, 47 and 57																								
Y	US 2010/0160833 (CHAU et al.) 24 June 2010 (24-06-2010) * Paragraphs [0010], [0030], [0039]-[0061], [0076] and [0077] *	12-14, 17-20, 31-33, 36-39 49, 59, 63 and 64																								
Y	US 2008/0269646 (CHAU et al.) 30 October 2008 * Paragraphs [0054]-[0056] and [0090] *	14, 15, 19, 20, 33, 34, 38 and 39																								
<p><input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.</p>		<p><input checked="" type="checkbox"/> See patent family annex.</p>																								
*	Special categories of cited documents :	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention																								
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone																								
"E"	earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art																								
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family																								
"O"	document referring to an oral disclosure, use, exhibition or other means																									
"P"	document published prior to the international filing date but later than the priority date claimed																									
<p>Date of the actual completion of the international search 21 December 2011 (21-12-2011)</p>		<p>Date of mailing of the international search report 9 January 2012 (09-01-2012)</p>																								
<p>Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476</p>		<p>Authorized officer Alan Chan (819) 934-7888</p>																								

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/CA2011/001085

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2005/0286772 (ALBERTELLI) 29 December 2005 (29-12-2005)	
A	US 2010/0250473 (PORIKLI et al.) 30 September 2010 (30-09-2010)	
A	US 2006/0074823 (HEUMANN et al.) 06 April 2006 (06-04-2006)	
A	US 2005/0283096 (CHAU et al.) 22 December 2005 (22-12-2005)	

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CA2011/001085

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
US2006120609A1	08 June 2006 (08-06-2006)	JP4719224B2 JP2008525869A US7742641B2 WO2006062863A2 WO2006062863A3	06 July 2011 (06-07-2011) 17 July 2008 (17-07-2008) 22 June 2010 (22-06-2010) 15 June 2006 (15-06-2006) 15 February 2007 (15-02-2007)
US2010160833A1	24 June 2010 (24-06-2010)	None	
US2008269646A1	30 October 2008 (30-10-2008)	AU2005253661A1 AU2005253661B2 CA2570202A1 CN101005799A EP1773184A1 EP1773184A4 JP4773432B2 JP2008502386A US2009227908A1 US7749177B2 US2005283096A1 WO2005122877A1	29 December 2005 (29-12-2005) 03 June 2010 (03-06-2010) 29 December 2005 (29-12-2005) 25 July 2007 (25-07-2007) 18 April 2007 (18-04-2007) 30 September 2009 (30-09-2009) 14 September 2011 (14-09-2011) 31 January 2008 (31-01-2008) 10 September 2009 (10-09-2009) 06 July 2010 (06-07-2010) 22 December 2005 (22-12-2005) 29 December 2005 (29-12-2005)
US2005286772A1	29 December 2005 (29-12-2005)	None	
US2010250473A1	30 September 2010 (30-09-2010)	JP2010231768A	14 October 2010 (14-10-2010)
US2006074823A1	06 April 2006 (06-04-2006)	CN1749987A SG121074A1 US7373332B2	22 March 2006 (22-03-2006) 26 April 2006 (26-04-2006) 13 May 2008 (13-05-2008)
US2005283096A1	22 December 2005 (22-12-2005)	AU2005253661A1 AU2005253661B2 CA2570202A1 CN101005799A EP1773184A1 EP1773184A4 JP4773432B2 JP2008502386A US2009227908A1 US7749177B2 US2008269646A1 WO2005122877A1	29 December 2005 (29-12-2005) 03 June 2010 (03-06-2010) 29 December 2005 (29-12-2005) 25 July 2007 (25-07-2007) 18 April 2007 (18-04-2007) 30 September 2009 (30-09-2009) 14 September 2011 (14-09-2011) 31 January 2008 (31-01-2008) 10 September 2009 (10-09-2009) 06 July 2010 (06-07-2010) 30 October 2008 (30-10-2008) 29 December 2005 (29-12-2005)