# 



[Information Fusion 16 (2014) 3–17](http://dx.doi.org/10.1016/j.inffus.2013.04.006)

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com/science/journal/15662535)

Information Fusion

journal homepage: [www.elsevier.com/locate/inffus](http://www.elsevier.com/locate/inffus)



A survey of multiple classiﬁer systems as hybrid systems



Michał Woz´niak [a](#_bookmark0),[⇑](#_bookmark3), Manuel Graña [b](#_bookmark1), Emilio Corchado [c](#_bookmark2)

a *Department of Systems and Computer Networks, Wroclaw University of Technology, Wroclaw, Poland*

b *Computational Intelligence Group, University of the Basque Country, San Sebastian, Spain*

c *Departamento de Informática y Automática, University of Salamanca, Salamanca, Spain*

a r t i c l e i n f o

*Article history:*

Available online 29 April 2013

*Keywords:*

Combined classiﬁer Multiple classiﬁer system Classiﬁer ensemble Classiﬁer fusion

Hybrid classiﬁer

a b s t r a c t

A current focus of intense research in pattern classiﬁcation is the combination of several classiﬁer sys- tems, which can be built following either the same or different models and/or datasets building approaches. These systems perform *information fusion* of classiﬁcation decisions at different levels over- coming limitations of traditional approaches based on single classiﬁers. This paper presents an up-to- date survey on multiple classiﬁer system (MCS) from the point of view of Hybrid Intelligent Systems. The article discusses major issues, such as diversity and decision fusion methods, providing a vision of the spectrum of applications that are currently being developed.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Hybrid Intelligent Systems offer many alternatives for unortho- dox handling of realistic increasingly complex problems, involving

ambiguity, uncertainty and high-dimensionality of data. They al- low to use both *a priori* knowledge and raw data to compose inno- vative solutions. Therefore, there is growing attention to this multidisciplinary research ﬁeld in the computer engineering re- search community. Hybridization appears in many domains of hu- man activity. It has an immediate natural inspiration in the human biological systems, such as the Central Nervous System, which is a *de facto* hybrid composition of many diverse computational units, as discussed since the early days of computer science, e.g., by von Neumann [[1]](#_bookmark12) or Newell [[2]](#_bookmark13). Hybrid approaches seek to exploit the strengths of the individual components, obtaining enhanced performance by their combination. The famous ‘‘*no free lunch*’’ the- orem [[3]](#_bookmark14) stated by Wolpert may be extrapolated to the point of saying that there is no single computational view that solves all problems. [Fig. 1](#_bookmark4) is a rough representation of the computational do- mains covered by the Hybrid Intelligent System approach. Some of them deal with the uncertainty and ambiguity in the data by prob- abilistic or fuzzy representations and feature extraction. Others deal with optimization problems appearing in many facets of the

\* Corresponding author.

*E-mail addresses:* [michal.wozniak@pwr.wroc.pl](mailto:michal.wozniak@pwr.wroc.pl) (M. Woz´niak), [ccpgrrom@g-](mailto:ccpgrrom@gmail.com) [mail.com](mailto:ccpgrrom@gmail.com) (M. Graña), [escorchado@usal.es](mailto:escorchado@usal.es) (E. Corchado).

intelligent system design and problem solving, either following a nature inspired or a stochastic process approach. Finally, classiﬁers implementing the intelligent decision process are also subject to hybridization by various forms of combination. In this paper, we focus in this speciﬁc domain, which is in an extraordinary efferves- cence nowadays, under the heading of Multi-Classiﬁer Systems (MCS). Referring to classiﬁcation problems, Wolpert’s theorem has an speciﬁc lecture: there is not a single classiﬁer modeling ap- proach which is optimal for all pattern recognition tasks, since each has its own domain of competence. For a given classiﬁcation task, we expect the MCS to exploit the strengths of the individual classiﬁer models at our disposal to produce the high quality com- pound recognition system overcoming the performance of individ- ual classiﬁers. Summarizing:

Hybrid Intelligent Systems (HIS) are free combinations of compu- tational intelligence techniques to solve a given problem, cover- ing al computational phases from data normalization up to ﬁnal decision making. Speciﬁcally, they mix heterogeneous funda- mental views blending them into one effective working system. Information Fusion covers the ways to combine information sources in a view providing new properties that may allow to solve better or more efﬁciently the proposed problem. Informa- tion sources can be the result of additional computational processes.

●

●

Multi-Classiﬁer Systems (MCS) focus on the combination of classiﬁers form heterogenous or homogeneous modeling back- grounds to give the ﬁnal decision. MCS are therefore a subcate- gory of HIS.

●

1566-2535/$ - see front matter © 2013 Elsevier B.V. All rights reserved. <http://dx.doi.org/10.1016/j.inffus.2013.04.006>

4 *M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17*

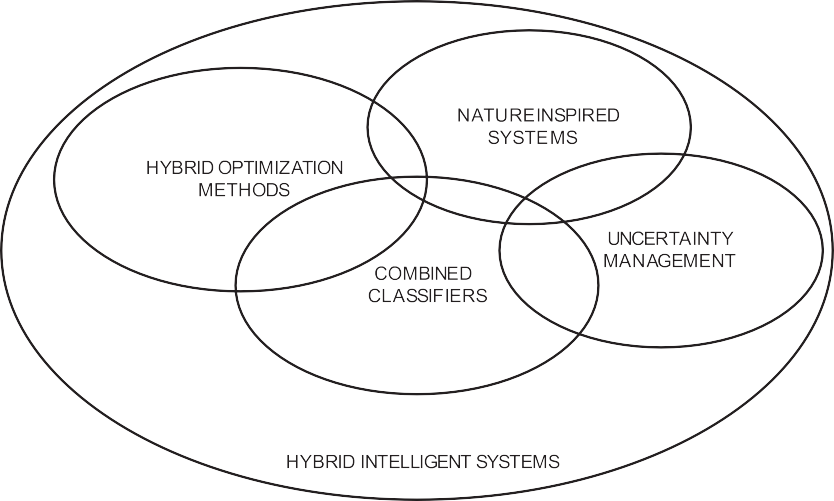


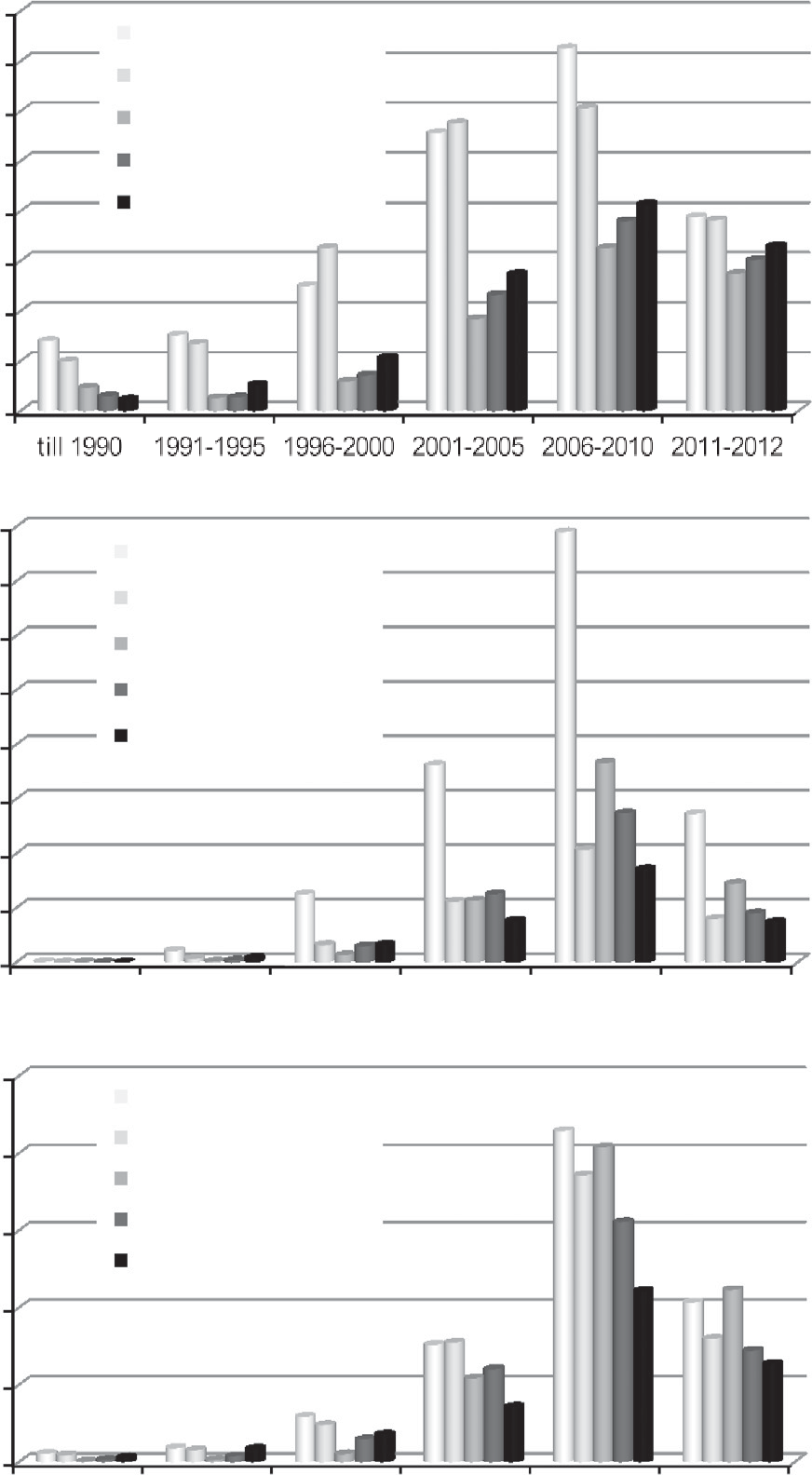
Fig. 1. Domains of hybrid intelligent systems.

*Historical perspective*. The concept of MCS was ﬁrst presented by Chow [[4]](#_bookmark15), who gave conditions for optimality of the joint decision[1](#_bookmark5) of independent binary classiﬁers with appropriately deﬁned weights. In 1979 Dasarathy and Sheela combined a linear classiﬁer and one *k*- NN classiﬁer [[6]](#_bookmark16), suggesting to identify the region of the feature space where the classiﬁers disagree. The *k*-NN classiﬁer gives the an- swer of the MCS for the objects coming from the conﬂictive region and by the linear one for the remaining objects. Such strategy signif- icantly decreases the exploitation cost of whole classiﬁer system. This was the ﬁrst work introducing a classiﬁer selection concept, however the same idea was developed independently in 1981 by Rastrigin and Erenstein [[7]](#_bookmark18) performing ﬁrst a feature space partition- ing and, second, assigning to each partition region an individual clas- siﬁer that achieves the best classiﬁcation accuracy over it. Other early relevant works formulated conclusions regarding MCS ’s classi- ﬁcation quality, such as [[8]](#_bookmark19) who considered a neural network ensem- ble, [[9]](#_bookmark20) with majority voting applied to handwriting recognition, Turner in 1996 [[10]](#_bookmark21) showed that averaging outputs of an inﬁnite number of unbiased and independent classiﬁers can lead to the same response as the optimal Bayes classiﬁer, Ho [[11]](#_bookmark22) underlined that a decision combination function must receive useful representation of each classiﬁer’s decision. Speciﬁcally, they considered several method based on decision ranks, such as Borda count. Finally, the landmark works devoted introducing bagging [[12]](#_bookmark23) and boosting [[13,14]](#_bookmark24) which are able to produce strong classiﬁers [[15]](#_bookmark25), in the (*Prob- ably Approximately Correct*) theory [[16]](#_bookmark26) sense, on the basis of the weak one. Nowadays MCS, are highlighted by review articles as a hot topic and promising trend in pattern recognition [[17–21]](#_bookmark28). These reviews include the books by Kuncheva [[22]](#_bookmark31), Rokach [[23]](#_bookmark32), Seni and Edler [[24]](#_bookmark33), and Baruque and Corchado [[25]](#_bookmark33). Even leading-edge gen- eral machine learning handbooks such as [[26–28]](#_bookmark33) include extensive presentations of MCS concepts and architectures. The popularity of this approach is conﬁrmed by the growing trend in the number of publications shown in [Fig. 2](#_bookmark6). The ﬁgure reproduces the evolution of the number of references retrieved by the application of speciﬁc keywords related to MCS since 1990. The experiment was repeated

on three well known academic search sites. The growth in the num- ber of publications has an exponential trend. The last entry of the

1 We can retrace decision combination long way back in history. Perhaps the ﬁrst worthy reference is the Greek democracy (meaning *government of the people*) ruling that full citizens have an equal say in any decision that affects their life. Greeks believed in the community wisdom, meaning that the rule of the majority will produce the optimal joint decision. In 1785 Condorcet formulated the *Jury Theorem* about the misclassiﬁcation probability of a group of independent voters [[5]](#_bookmark17)], providing the ﬁrst result measuring the quality of classiﬁer committee.

Fig. 2. Evolution of the number of publications *per* year ranges retrieved from the keywords speciﬁed in the plot legend. Each plot corresponds to searching site: the top to Google Scholar; the center to the Web of Knowledge, the bottom to Scopus. The ﬁrst entry of the plots is for publications prior to 1990. The last entry is only for the last 2 years.



*M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17* 5

plots corresponds to the last 2 years, and some of the keywords give as many references as in the previous 5 years.

*Advantages*. Dietterich [[29]](#_bookmark33) summarized the beneﬁts of MCS: (a) allowing to ﬁlter out hypothesis that, though accurate, might be incorrect due to a small training set, (b) combining classiﬁers trained starting from different initial conditions could overcome the local optima problem, and (c) the true function may be impos- sible to be modeled by any single hypothesis, but combinations of hypotheses may expand the space of representable functions. Rephrasing it, there is widespread acknowledgment of the follow- ing advantages of MCS:

MCS behave well in the two extreme cases of data availability: when we have very scarce data samples for learning, and when we have a huge amount of them at our disposal. In the scarcity case, MCS can exploit bootstrapping methods, such as bagging

●

or boosting. Intuitive reasoning justiﬁes that the worst classiﬁer would be out of the selection by this method [[30]](#_bookmark33), e.g., by indi- vidual classiﬁer output averaging [[31]](#_bookmark33). In the event of availabil-

ity of a huge amount of learning data samples, MCS allow to

train classiﬁers on dataset’s partitions and merge their decision

using appropriate combination rule [[20]](#_bookmark29).

Combined classiﬁer can outperform the best individual classi- ﬁer [[32]](#_bookmark33). Under some conditions (e.g., majority voting by a group of independent classiﬁers) this improvement has been proven analytically [[10]](#_bookmark21).

●

Many machine learning algorithms are *de facto* heuristic search algorithms. For example the popular decision tree induction method C4.5 [[33]](#_bookmark33) uses a greedy search approach, choosing the search direction according to an heuristic attribute evaluation function. Such an approach does not assure an optimal solution. Thus, the combined algorithm, which could start its work from different initial points of the search space, is equivalent to a multi-start local random search which increases the probability of ﬁnding an optimal model.

●

MCS can easily be implemented in efﬁcient computing environ- ments such as parallel and multithreaded computer architec- tures [[34]](#_bookmark33). Another attractive area of implementation solutions is distributed computing systems (i.e.: P2P, Grid or Cloud computing) [[35,36]](#_bookmark33), especially when a database is parti- tioned for privacy reasons [[37]](#_bookmark34) so that partial solutions must be computed on each partition and only the ﬁnal decision is available as the combination of the networked decision.

●

Wolpert stated that each classiﬁer has its speciﬁc competence domain [[3]](#_bookmark14), where they overcome other competing algorithms, thus it is not possible to design a single classiﬁer which outper- forms another ones for each classiﬁcation tasks. MCS try to select always the local optimal model from the available pool of trained classiﬁers.

●

*System structure*. The general structure of MCS is depicted in [Fig. 3](#_bookmark7) following a classical pattern recognition [[38]](#_bookmark35) application structure. The most informative or discriminant features describ- ing the objects are input to the classiﬁer ensemble, formed by a set of *complementary* and *diverse* classiﬁers. An appropriate fusion method combines the individual classiﬁer outputs optimally to provide the system decision. According to Ho [[39]](#_bookmark39), two main MCS design approaches can be distinguished. On one hand, the so-called *coverage optimization* approach tries to cover the space of possible models by the generation of a set of mutually comple- mentary classiﬁers whose combination provides optimal accuracy. On the other hand, the so-called *decision optimization* approach concentrates on designing and training an appropriate decision combination function over a set of individual classiﬁer given in ad- vance [[40]](#_bookmark42).The main issues in MCS design are:

System topology: How to interconnect individual classiﬁers. Ensemble design: How to drive the generation and selection of a pool of valuable classiﬁers.

●

●

Fuser design: How to build a decision combination function (fuser) which can exploit the strengths of the selected classiﬁers and combine them optimally.

●

1. System topology

[Fig. 4](#_bookmark8) illustrates the two canonical topologies employed in MCS design. The overwhelming majority of MCS reported in the litera- ture is structured in a parallel topology [[22]](#_bookmark31). In this architecture, each classiﬁer is feed the same input data, so that the ﬁnal decision of the combined classiﬁer output is made on the basis of the out- puts of the individual classiﬁers obtained independently. Alterna- tively, in the serial (or conditional) topology, individual classiﬁers are applied in sequence, implying some kind of ranking or ordering over them. When the primary classiﬁer cannot be trusted to clas- sify a given object, e.g., because of the low support/conﬁdence in its result, then the data is feed to a secondary classiﬁer [[41,42]](#_bookmark43), and so on, adding classiﬁers in sequence. This topology is adequate when the cost of classiﬁer exploitation is important, so that the pri- mary classiﬁer is the computationally cheapest one, and secondary classiﬁers have higher exploitation cost [[43]](#_bookmark45). This model can be ap- plied to classiﬁers with the so-called reject option as well [[44]](#_bookmark48). In

[[45]](#_bookmark50) the ﬁrst classiﬁer in the pipeline gives an estimation of the

certainty of the classiﬁcation, so that uncertain data samples are sent to a second classiﬁer, specialized in difﬁcult instances. We no- tice the similarity of such approach to the ordered set of rules [[46]](#_bookmark52) or decision list [[47]](#_bookmark53), when we consider each rule as the classiﬁer. A very special case of sequential topology is the Adaboost intro- duced by Freund and Schapire in 1995 [[48]](#_bookmark54), widely applied in data

***acer***

*2018-05-28 00:15:19*

--------------------------------------------

note

# 

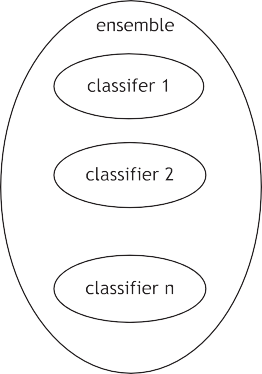
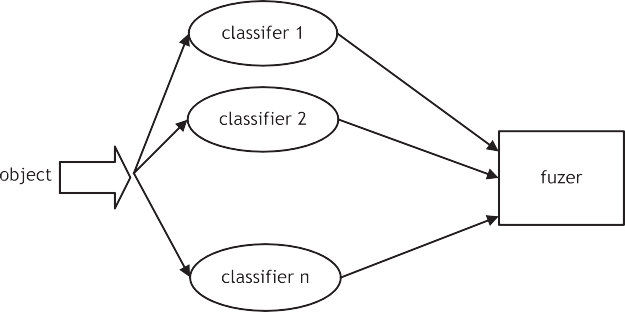


Fig. 3. Overview of multiple classiﬁer system.

6 *M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17*





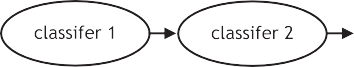


Fig. 4. The canonical topologies of MCSs: parallel (top) and serial (bottom).

mining problems [[49]](#_bookmark55). The goal of boosting is to enhance the accu- racy of any given learning algorithm, even weak learning algo- rithms with an accuracy slightly better than chance. Shapire [[50]](#_bookmark56) showed that weak learners can be *boosted* into a strong learning algorithm by sequentially focusing on the subset of the training data that is *hardest* to classify. The algorithms performs training of the weak learner multiple times, each time presenting it with an updated distribution over the training examples. The distribu- tion is altered so that *hard* parts of the feature space have higher probability, i.e. trying to achieve a hard margin distribution. The decisions generated by the weak learners are combined into a ﬁnal single decision. The novelty of Adaboost lies in the adaptability of the successive distributions to the results of the previous weak learners, thus the name AdaptiveBoost. In the words of Kivinen et al. [[51]](#_bookmark58), AdaBoost ﬁnds a new distribution that is closest to the old one but taking into consideration the restriction that the new distribution must be orthogonal to the mistake vector of the cur- rent weak learner.

1. Ensemble design

Viewing MCS as a case of robust software [[52–55]](#_bookmark59), diversity arises as the guiding measure of the design process. Classiﬁer ensemble design aims to include mutually complementary individ- ual classiﬁers which are characterized by high diversity and accu- racy [[56]](#_bookmark64). The emphasis from the Hybrid Intelligent System point of view is in building MCS from components following different kinds of modeling and learning approaches, expecting an increase in diversity and a decrease in classiﬁer output correlation [[57]](#_bookmark36). Unfortunately, the problem of how to measure classiﬁer diversity is still an open research topic. Brown et al. [[58]](#_bookmark36) notice that we can ensure diversity using implicit or explicit approaches. Implicit approaches include techniques of independent generation of indi- vidual classiﬁers, often based on random techniques, while explicit approaches focus on the optimization of a diversity metric over a given ensemble line-up. In this second kind of approaches, individ- ual classiﬁer training is performed conditional to the previous clas- siﬁers with the aim of exploiting the strengths of valuable members of classiﬁer pool. This section discusses some diversity measures, and the procedures followed to ensure diversity in the ensemble.

* 1. *Diversity measures*

For regression problems, the variance of the outputs of ensem- ble members is a convenient diversity measure, because it was

proved that the error of a compound model based on a weighted averaging of individual model outputs can be reduced according to increasing diversity [[56,59]](#_bookmark64). Brown et al. [[60]](#_bookmark36) showed a func- tional relation between diversity and individual regressor accu- racy, allowing to control the bias-variance tradeoff systematically. For classiﬁcation problems such theoretical results have not been proved yet, however many diversity measures have been pro- posed till now. On the one hand, it is intuitive that increasing diversity should lead to the better accuracy of the combined sys- tem, but there is no formal proof of this dependency [[61]](#_bookmark36), as con- ﬁrmed by the wide range of experimental results presented, e.g., in [[62]](#_bookmark36). In [[53]](#_bookmark60) authors decomposed the error of the classiﬁcation by majority voting into individual accuracy, *good* and *bad* diversi- ties. The *good* diversity has positive impact on ensemble error reduction, whereas the *bad* diversity has the opposite effect. Shark- ley et al. [[55]](#_bookmark62) proposed a hierarchy of four levels of diversity according to the answer of the majority rule, coincident failures, and possibility of at least one correct answer of ensemble mem- bers. Brown et al. [[58]](#_bookmark36) argue that this hierarchy is not appropriate when the ensemble diversity varies between feature subspaces.

They formulated the following taxonomy of diversity measures:

Pairwise measures averaging a measure between each classiﬁer pair in an ensemble, such as Q-statistic [[58]](#_bookmark36), kappa-statistics [[63]](#_bookmark36), disagreement [[64]](#_bookmark36) and double-fault measure [[61,65]](#_bookmark36).

●

Non-pairwise diversity measures comparing outputs of a given classiﬁer and the entire ensemble, such as Kohavi–Wolpert var- iance [[66]](#_bookmark36), a measure of inter-rater (inter-classiﬁer) reliability [[67]](#_bookmark36), the entropy measure [[68]](#_bookmark37), the measure of difﬁculty [[8]](#_bookmark19), generalized diversity [[52]](#_bookmark59), and coincident failure diversity [[69]](#_bookmark38).

●

The analysis of several diversity measures [[70]](#_bookmark40) relating them to the concept of classiﬁers’ margin, showed their limitations and the source of confusing empirical results. They relate the classiﬁer selec- tion to a NP-complete matrix cover problem, implying that ensem- ble design in fact a quite difﬁcult combinatorial problem. Diversity measures usually employ the most valuable sub-ensemble in *ensemble pruning* processes [[71]](#_bookmark41). To deal with the high computa- tional complexity of ensemble pruning, several hybrid approaches have been proposed such as heuristic techniques [[72,73]](#_bookmark44), evolution- ary algorithms [[74,75]](#_bookmark46), reinforcement learning [[76]](#_bookmark47), and competi- tive cross-validation techniques [[77]](#_bookmark49). For classiﬁcation tasks, the cost of acquiring feature values (which could be interpreted as the price for examination or time required to collect the data for deci- sion making) can be critical. Some authors take it into consideration during the component classiﬁer selection step [[78,79]](#_bookmark51).

*M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17* 7

* 1. *Ensuring diversity*

According to [[22,38]](#_bookmark31) we can enforce the diversity of a classiﬁer pool by the manipulation of either individual classiﬁer inputs, out- puts, or models.

* + 1. *Diversifying input data*

This diversiﬁcation strategy assumes that classiﬁers trained on different (disjoint) input subspaces become complementary. Three general strategies are identiﬁed:

1. Using different data partitions.
2. Using different sets of features.
3. Taking into consideration the local specialization of individual classiﬁers.

*Data partitions* They may be compelled by several reasons, such as data privacy, or the need to learn over distributed data chunks stored in different databases [[80–82]](#_bookmark57). Regarding data privacy, we should notice that using distributed data may come up against le- gal or commercial constraints which do not allow sharing raw datasets and merging them into a common repository [[37]](#_bookmark34). To en- sure privacy we can train individual classiﬁers on each database independently and merge their outputs using hybrid classiﬁer principles [[83]](#_bookmark61). The distributed data paradigm is strongly con- nected with the big data analysis problem [[84]](#_bookmark63). A huge database may impede to deliver trained classiﬁers under speciﬁed time con- straints, imposing to resort to sampling techniques to obtain man- ageable dataset partitions. A well known approach is cross- validated committee which requires to minimize overlapping of dataset partitions [[56]](#_bookmark64). Providing individualized train datasets for each classiﬁer is convenient in the case of shortage of learning examples. Most popular techniques, such as bagging [[12]](#_bookmark23) or boost- ing [[14,19,64,85]](#_bookmark27), have their origin in bootstrapping [[13]](#_bookmark24). These methods try to ascertain if a set of weak classiﬁer may produce a strong one. Bagging applies sampling with replacement to obtain independent training datasets for each individual classiﬁer. Boost- ing modiﬁes the input data distribution perceived by each classiﬁer from the results of classiﬁers trained before, focusing on difﬁcult samples, making the ﬁnal decision by a weighted voting rule.

*Data features* May be selected to ensure diversity training of a

pool of classiﬁers. The Random Subspace [[86,87]](#_bookmark65) was employed for several types of the individual classiﬁers such as decision tree (Random Forest) [[88]](#_bookmark65), linear classiﬁers [[89]](#_bookmark65), or minimal distance classiﬁer [[90,91]](#_bookmark65). It is worth pointing out the interesting proposi- tions dedicated one-class classiﬁer presented by Nanni [[92]](#_bookmark65) or an hierarchical method of ensemble forming, based on feature space splitting and then assigning two-class classiﬁers (i.e. Support Vec- tor Machines) locally presented in [[93,94]](#_bookmark65). Attribute Bagging [[95]](#_bookmark66) is a wrapper method that establishes the appropriate size of a feature subset, and then creates random projections of a given training set by random selection of feature subsets. The classiﬁer ensemble are train on the basis of the obtained set.

*Local specialization* It is assumed for classiﬁer selection, select-

ing the best single classiﬁer from a pool of classiﬁers trained over each partition of the feature space. It gives the MCS answer for all objects included in the partition [[7]](#_bookmark18). Some proposals assume clas- siﬁer local specialization, providing only locally optimal solutions [[38,96–98,72]](#_bookmark35), while others divide the feature space, selecting (or training) a classiﬁer for each partition. Static and dynamic ap- proaches are distinguished:

Static classiﬁer selection [[99]](#_bookmark72): the relation between region of competence and assigned classiﬁer is ﬁxed. Kuncheva’s Cluster- ing and Selection algorithm [[100]](#_bookmark74) partitions the feature space by a clustering algorithm, and selects the best individual classi-

●

ﬁer for each cluster according to its local accuracy. Adaptive Splitting and Selection algorithm in [[101]](#_bookmark77) partitions the feature space and assigns classiﬁers to each partition into one inte- grated process. The main advantage of AdaSS is that the training algorithm considers an area contour to determine the classiﬁer content and, conversely, that the region shapes adapt to the competencies of the classiﬁers. Additionally, the majority vot- ing or more sophisticated rules are proposed as combination method of area classiﬁers [[102]](#_bookmark78). Lee et al. [[103]](#_bookmark82) used the fuzzy entropy measure to partition the feature space and select the relevant features with good separability for each of them.

Dynamic classiﬁer selection: the competencies of the individual

●

classiﬁers are calculated during classiﬁcation operation [[104–](#_bookmark84) [107]](#_bookmark84). There are several interesting proposals which extend this concept, e.g., by using preselected committee of the individual classiﬁer and making the ﬁnal decision on the basis of a voting rule [[108]](#_bookmark86). In [[109,110]](#_bookmark87) authors propose dynamic ensemble selection based on the original competence measure using clas- siﬁcation of so-called random reference classiﬁer.

Both static [[111–113]](#_bookmark89) and dynamic [[114–116]](#_bookmark67) classiﬁer special- ization are widely used for data stream classiﬁcation.

* + 1. *Diversifying outputs*

MCS diversity can be enforced by the manipulation of the indi- vidual classiﬁer outputs, so that an individual classiﬁer is designed to classify only some classes in the problem.

The combination method should restore the whole class label set, e.g., a multi-class classiﬁcation problem can be decomposed into a set of binary classiﬁcation problems [[117,118]](#_bookmark67). The most popular propositions of two-class classiﬁer combinations are:

OAO (*one-against-one*) and OAA (*one-against-all*)[[119]](#_bookmark67), where at least one predictor relates to each class. The model that a given ob- ject belongs to a chosen class is tested against the alternative of the feature vector belonging to any other class. In the OAA method, a classiﬁer is trained to separate a chosen class from the remaining ones. OAA returns class with maximum support. In more general approaches, the combination of individual outputs is made by ﬁnd- ing the closest class, in some sense, to the code given by the out- puts of the individual classiﬁers. ECOC (*Error Correcting Output Codes*) model was proposed by Dieterich and Bakiri [[118]](#_bookmark67), who as- sumed that a set of classiﬁers produces sequence of bits which is related to code-words during training. The ECOC points at the class with the smallest Hamming distance to its codeword. Passerini et al. showed advantages of this method over traditional ones for the ensemble of support vector machines [[120]](#_bookmark67).

Recently several interesting propositions on how to combine

the binary classiﬁers were proposed. Wu et al. [[121]](#_bookmark67) used pairwise coupling, Friedman employed Max-Win rule [[122]](#_bookmark68), Hüllermeier proposed the adaptive weighted voting procedure [[123]](#_bookmark69). A com- prehensive recent survey of binary classiﬁer ensembles is [[124]](#_bookmark70). It worth mentioning the one-class classiﬁcation model which is the special case of binary classiﬁer trained in the absence of coun- terexamples. Its main goal is to model normality in order to detect anomaly or outliers from the target class [[125]](#_bookmark71). To combine such classiﬁers the typical methods developed for binary ones are used

[[126]](#_bookmark73) but it is worth mention the work by Wilk and Wozniak where authors restored multi-class classiﬁcation task using a pool of one-class classiﬁers and the fuzzy inference system [[127]](#_bookmark75). The combination methods dedicated the one-class classiﬁers still await a proper attention [[128]](#_bookmark76).

* + 1. *Diversifying models*

Ensembles with individual classiﬁers based on different classiﬁ- cation models take advantage of the different biases of each classi- ﬁer model [[3]](#_bookmark14). However, the combination rule should be carefully

8 *M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17*

chosen. We can combine the class labels but in the case of contin- uous outputs we have to normalize them, e.g., using fuzzy ap- proach [[127]](#_bookmark75). We could use the different versions of the same model as well, because many machine learning algorithms do not guarantee to ﬁnd the optimal classiﬁer. Combining the results of various initializations may give good results. Alternatively, a pool of classiﬁers can be produced by noise injection. Regarding neural networks [[129]](#_bookmark79) it is easy to train pools of networks where each of them is trained starting from randomly chosen initial weights. Regarding decisions tree we can choose randomly the test for a gi- ven node among the possible tests according to the value of a split- ting criterion.

1. Fuser design

Some works consider the answers from a given Oracle as the reference combination model [[130]](#_bookmark80). The Oracle is an abstract com- bination model, built such that if at least one of the individual clas- siﬁers provides the correct answer, then the MCS committee outputs the correct class too. Some researches used the Oracle in comparative experiments to provide a performance upper bound for classiﬁer committee [[10]](#_bookmark21) or information fusion methods [[131]](#_bookmark81). A simple example shows the risks of the Oracle model: as- sume we have two classiﬁers for a binary class problem, a random one and the other that always returns the opposite decision; hence the Oracle will always return the correct answer. As a consequence the Oracle model does not ﬁt in the Bayesian paradigm. Raudys

[[132]](#_bookmark83) noticed that Oracle is a kind of quality measure of a given individual classiﬁer pool. Let us systematize methods of classiﬁer fusion, which on the one hand could use class labels or support function, on the other hand combination rules could be given or be the results of training. The taxonomy of decision fusion strate- gies is depicted in [Fig. 5](#_bookmark9).

* 1. *Class label fusion*

Early algorithms performing fusion of classiﬁer responses [[9,10,61]](#_bookmark20) only implemented majority voting schemes in three main versions [[22]](#_bookmark31):



Fig. 5. A taxonomy of fusing strategies for the combination of MCS individual decisions.

unanimous voting, so that the answer requires that all classiﬁ- ers agree,

simple majority, so that the answer is given if majority is greater than half the pool of classiﬁers,

●

●

majority voting, taking the answer with the highest number of votes.

●

The expected error of majority voting (for independent classiﬁ- ers with the same quality) was estimated in 1794 according to Ber- noulli’s equation, proven as the Condorcet Jury Theorem [[5]](#_bookmark17). Later works focused on the analytically derived classiﬁcation perfor- mance of combined classiﬁers hold only when strong conditions are met [[8]](#_bookmark19) so that they are not useful from practical point of view. Alternative voting methods weight differently the decisions com- ing from different committee members [[22,133]](#_bookmark31). The typical archi- tecture of combined classiﬁer based on class labels is presented in the left diagram of [Fig. 6](#_bookmark10). In [[134]](#_bookmark85) authors distinguished the types of weighted voting according to the classiﬁer, both to the classiﬁer and the class, and, ﬁnally, to features values, the classiﬁer and the class. Anyway, no one of these models can improve over the Oracle. To achieve that we need additional information, such as the feature values [[132,135,136]](#_bookmark83) as depicted in the right diagram of [Fig. 6](#_bookmark10).

* 1. *Support function fusion*

Support function fusion system architecture is depicted in [Fig. 7](#_bookmark11). Support functions provide a score for the decision taken by an individual classiﬁer. The value of a support function is the estimated likelihood of a class, computed either as a neural net- work output, *a posteriori* probability, or fuzzy membership func- tion. First to be mentioned, the Borda count [[11]](#_bookmark22) computes an score for each class on the basis of its ranking by each individual classiﬁer. The most popular form of support function is the *a pos- teriori* probability [[26]](#_bookmark33), produced by the probabilistic models embodied by the classiﬁers [[137–139]](#_bookmark88). There are many works fol- lowing this approach, such as the optimal projective fuser of [[140]](#_bookmark90), the combination of neural networks outputs according to their accuracy [[141]](#_bookmark91), and Naïve Bayes as the MCS combination method [[142]](#_bookmark92).

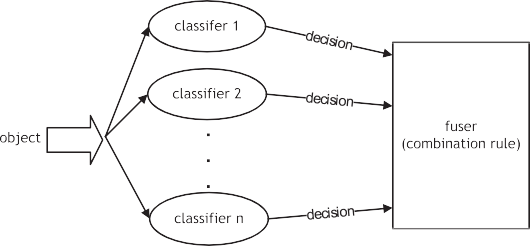
Some analytical properties and experimental evaluations of

aggregating methods were presented in [[10,31,143,144]](#_bookmark21). The aggregating methods use simple operators such as supremum or the mean value. They do not involve learning. However, they have little practical applicability because of the hard conditions imposed by them [[145]](#_bookmark92). The main aggregating advantage is that it counter- acts over-ﬁtting of individual classiﬁers. According to [[134]](#_bookmark85), the following types of weighted aggregation can be identiﬁed depend- ing on: (a) only the classiﬁer id, (b) the classiﬁer and the feature vector, (c) on the classiﬁer and the class, and (d) on the classiﬁer, the class, and the feature vector. For two-class recognition prob- lems only the last two types of aggregation allow to produce com- pound classiﬁer which may improve the Oracle. For many-class problems, it is possible to improve the Oracle [[131]](#_bookmark81) using any of these aggregation methods. Finally, another salient approach is the *mixture of experts* [[146,147]](#_bookmark92) which combines classiﬁer outputs using so-called input dependent gating function. Tresp and Tanig- uchi [[148]](#_bookmark92) proposed a linear function for this fuser model, and Cheeseman [[149]](#_bookmark92) proposed a mixture of Gaussian.

* 1. *Trainable Fuser*

Fuser weight selection can be treated as a speciﬁc learning pro- cess [[31,136]](#_bookmark33). Shlien [[150]](#_bookmark92) used Dempster and Shafer’s theory to reach a consensus on the weights to combine decision trees. Woz- niak [[151]](#_bookmark92) trained the fuser using perceptron-like learning, evolu- tionary algorithm [[152,153]](#_bookmark93). Zheng used data envelopment

*M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17* 9



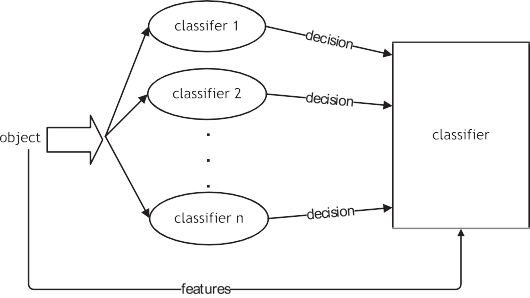


Fig. 6. Architecture of the MCS making decision on the basis of class label fusion only (left diagram). The right diagram corresponds to a MCS using additional information from the feature values.

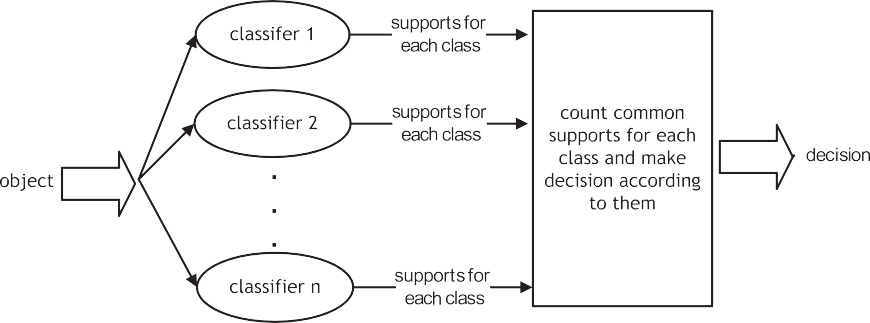


Fig. 7. Architecture of the MCS which computes the decision on the basis of support function combination.

analysis [[154]](#_bookmark97). Other fuser trainable methods may be strictly re- lated to ensemble pruning methods, when authors use some heu- ristic search algorithm to select the classiﬁer ensemble, as [[72,141]](#_bookmark44) according to the chosen fuser.

We have to mention the group of combination methods built from pools of heterogenous classiﬁers, i.e. using different classiﬁca- tion models, such as stacking [[155]](#_bookmark98). This method trains combina- tion block using individual classiﬁer outputs presented during classiﬁcation of the whole training set. Most of the combination methods do not take into consideration possible relations among individual classiﬁers. Huang and Suen [[156]](#_bookmark100) proposed Behavior- Knowledge Space method which aggregates the individual classiﬁ- ers decision on the basis of the statistical approach.

1. Concept Drift

Before entering the discussion of practical applications we con- sider a very speciﬁc topic of real life relevance which is known as Concept Drift in knowledge engineering domains, or non-station- ary processes in signal processing and statistics domains. Most of the conventional classiﬁers do not take into consideration this phe- nomenon. Concept Drift means that the statistical dependencies between object features and its classiﬁcation may change in time, so that future data may be badly processed if we maintain the same classiﬁcation, because the object category or its properties will be changing. Concept drift occurs frequently in real life [[157]](#_bookmark101). MCS are specially well suited to deal with Concept Drift.

Machine learning methods in security applications (like spam ﬁlters or IDS/IPS) [[158]](#_bookmark103) or decision support systems for marketing departments [[159]](#_bookmark105) require to take into account new training data with potentially different statistical properties [[116]](#_bookmark67). The occur- rence of Concept Drift decreases the true classiﬁcation accuracy dramatically. The most popular approaches are the Streaming Ensemble Algorithm (SEA) [[111]](#_bookmark89) and the Accuracy Weighted Ensemble (AWE) [[160]](#_bookmark107). Incoming data are collected in data chunks,

which are used to train new models. The individual classiﬁers eval- uation is done on their accuracy on the new data. The best per- forming classiﬁers are selected to constitute the MCS committee in the next time epoch. As the decision rule, the SEA uses a major- ity voting, whereas the AWE uses a weighted voting strategy. Ko- tler et al. present the Dynamic Weighted Majority (DWM) algorithm [[114]](#_bookmark67) which modiﬁes the decision combination weights and updates the ensemble according to number of incorrect deci- sions made by individual classiﬁers. When a classiﬁer weight is too small, then it is removed from the ensemble, a new classiﬁer is trained and added to the ensemble in its place.

A difﬁcult problem is drift detection, which is the problem of deciding that the Concept Drift has taken place. The current re- search direction is to propose an additional binary classiﬁer giv- ing the decision to rebuild the classiﬁers. The drift detector can be based on changes in the probability distribution of the in- stances [[161–163]](#_bookmark109) or classiﬁcation accuracy [[164,165]](#_bookmark113). Not all classiﬁcation algorithms dealing with concept drift require drift detection, because they can adjust the model to incoming data [[166]](#_bookmark116)[?].

1. Applications

Reported applications of classiﬁer ensembles have grown astoundingly in the recent years due to the increase in computa- tional power allowing training of large collections of classiﬁers in practical application time constraints. A recent review appears in [[18]](#_bookmark30). Sometimes the works combine diverse kinds of classiﬁers, so-called heterogeneous MCS. Homogeneous MCS, such as Random

Forest (RF), are composed of classiﬁers of the same kind. In the works revised below, basic classiﬁers are Multi-Layer Perceptron

(MLP), k-Nearest Neighbor (kNN), Radial Basis Function (RBF), Sup-

port Vector Machines (SVM), Probabilistic Neural Networks

(PNNs), and Maximum Likelihood (ML) classiﬁers.

10 *M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17*

We review in this section recent applications to remote sensing data, computer security, ﬁnancial risk assessment, fraud detection, recommender systems, and medical computer aided diagnosis.

* 1. *Remote sensing*

The main problems addressed by MCS in remote sensing do- mains are the land cover mapping and change detection. Land cov- er mapping consists in the identiﬁcation of materials that are in the surface of the area being covered. Depending on the application, a few general classes may be identiﬁed, i.e. vegetation, water, build- ings, roads, or a more precise classiﬁcation can be required, i.e. identifying tree or crop types. Applications include agriculture, for- estry, geology, urban planning, infrastructure degradation assess- ment. Change detection consists in the identiﬁcation of places where the land cover has changed in time, it implies the computa- tion over time series of images. Change detection may or may not be based on previous or separate land cover maps. Remote sensing classiﬁcation can be done on a variety of data sources, sometimes performing fusion of different data modalities. Optical data has better interpretability by humans, but land is easily occluded by weather conditions, i.e. cloud formations. Hyperspectral sensing provides high-dimensional data at each image pixel, with high spectral resolution. Synthetic Aperture Radar (SAR) is not affected by weather or other atmospheric conditions, so that observations are better suited for continuous monitoring of seasonally changing land covers. SAR can provide also multivariate data from varying radar frequencies. Other data sources are elevation maps, and other ancillary information, such as the measurements of environ- mental sensors.

* + 1. *Land cover mapping*

Early application of MCS to land cover mapping consisted in overproducing a large set of classiﬁers and searching for the opti- mal subset [[38,65,167]](#_bookmark35). To avoid the combinatorial complexity, the approach performs clustering of classiﬁer error, aggregating similar classiﬁers. The approach was proven to be optimal under some conditions on the classiﬁers. Interestingly, testing was per- formed on multi-source data, composing the pixel’s feature vector of joining multi-spectral with radar data channels, to compute the land cover map. The MCS was heterogenous, composed of MLP, RBF, and PNN.

The application of RF to processing remote sensing data has been abundant in the literature. It has been applied to estimate land cover on Landsat data over Granada, Spain [[168]](#_bookmark120) and multi- source data in a Colorado mountainous area [[169]](#_bookmark123). Speciﬁcally, Landsat Multi-Spectral, elevation, slope and aspect data are used as input features. The RF approach is able to successfully fuse these inhomogeneous informations. Works on hyperspectral images ac- quired by the HyMap sensor have been addressed to build vegeta- tion thematic maps [[170]](#_bookmark125), comparing RF and decision tree-based Adaboost, as well as two feature selection methods: the out-of- bag and a best-ﬁrst search wrapper feature subset selection meth- od. Diverse feature subsets are tested, and the general conclusion is that tree ecotopes are better discriminated than grass ecotopes. Further work with RF has been done assessing the uncertainty in modeling the distribution of vegetation types [[171]](#_bookmark94), performing classiﬁcation on the basis of environmental variables, in an ap- proach that combines spatial distribution modeling by spatial interpolation, using sequential Gaussian simulation and the clus- tering of species into vegetation types. Dealing with labeled data scarcity, there are methods [[172]](#_bookmark94) based on the combination of RF and the enrichment of the training dataset with artiﬁcially gener- ated samples in order to increase classiﬁer diversity, which is ap- plied to Landsat multispectral data. Artiﬁcial data is generated from the Gaussian modeling of the data distribution. The applica-

tion of RF to SAR multitemporal data aims to achieve season invari- ant detection of several classes of land cover, i.e. grassland, ceral, forest, etc. [[173]](#_bookmark94). RF performed best, with lowest spatial variability. Images were coregistered and some model portability was tested, where the model trained on one SAR image was applied on other SAR images of the same site obtained at different times. The suc- cess of RF for remote sensing images has prompted the proposal of an speciﬁc computational environment [[174]](#_bookmark94).

Ensambles of SVM have been also applied to land cover map. In- deed, the ground truth data scarcity has been attacked by an active learning approach to semi-supervised SVM training [[175]](#_bookmark94). The ac- tive learning approach is based on the clustering of the unlabeled data samples according to the clustering of the SVM outputs on the current training dataset. Samples with higher membership coefﬁcient are added to the corresponding class data, and the clas- siﬁer is retrained in an iterative process. These semi-supervised SVM are combined in a majority voting ensemble and applied to the classiﬁcation SPOT and Landsat optical data. Land cover classi- ﬁcation in the speciﬁc context of shallow waters has the additional difﬁculties of the scattering, refraction and reﬂection effects intro- duced by the water cover. A robust process combines a parallel and a serial architecture [[176]](#_bookmark94), where initial classiﬁcation results ob- tained by SVM are reﬁned in a second SVM classiﬁer and the ﬁnal result is given by a linear combination of two ensembles of SVM classiﬁers and a minimum distance classiﬁer. Besides, the system estimates the water depth by a bathymetry estimation process. The approach is applied to Landsat images for the estimation of coral population in coastal waters. Polarimetric SAR data used for the classiﬁcation of Boreal forests require an ensemble of SVM [[177]](#_bookmark94). Each of the SVM is speciﬁcally tuned to a class, with speciﬁc feature selection process. Best results are obtained when multi- temporal data is used, joining two images from two different sea- sons (summer and winter) and performing the feature selection and training on the joint data vectors.

* + 1. *Change detection*

Early application of MCS to land cover change detection was based on non-parametric algorithms, speciﬁcally MLP, k-NN, RBF, and ML classiﬁers [[178,179]](#_bookmark94), where classiﬁer fusion was performed either by majority voting, Bayesian average and maximum *a poste- riori* probability. Testing data were Thematic Mapper multispectral images, and the Synthetic Aperture Radar (SAR) of Landsat 5 satel- lite. Recent works on change detection in panchromatic images with MCS follow three different decision fuser strategies: majority voting, Dempster-Shafer evidence theory, and the Fuzzy Integral [[180]](#_bookmark95). The sequential process of the images previous to classiﬁca- tion includes pan-sharpening of the multi-temporal images, co- registration, raw radiometric change detection by image subtrac- tion and automatic thresholding, and a ﬁnal MCS decision com- puted on the multi-spectral data and the change detection data obtained from the various pan-sharpening approaches.

* 1. *Computer security*

Computer security is at the core of most critical services nowa- days, from universities, banking, companies, communication. Se- cure information processing is a growing concern, and the machine learning approaches are trying to provide predictive solu- tions that may allow to avoid the negative impact of such attacks. Here we introduce some of the problems, with current solutions proposed from the MCS paradigm.

* + 1. *Distributed denial of service*

Distributed denial of service (DDoS) are among the most threat- ening attacks that an Internet Service Provider may face. Distrib- uted service providers, such as military applications, e-healthcare

*M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17* 11

and e-governance can be very sensitive to this type of attacks, which can produce network performance degradation, service unavailability, and revenue loss. There is a need for intelligent sys- tems able to discriminate legitimate ﬂash crowds from an attack. A general architecture for automatic detection of DDoS attacks is needed where the attack detection may be performed by a MCS. The MCS constituent classiﬁers may be ANNs trained with robust learning algorithms, i.e. Resilient Back Propagation (RBP). Speciﬁ- cally, a boosting strategy is deﬁned on the ensemble of RBP trained ANNs, and a Neyman Pearson approach is used to make the ﬁnal decision [[181]](#_bookmark96). This architecture may be based on Sugeno Adaptive Neuro-Fuzzy Inference Systems (ANFIS) [[182]](#_bookmark99). A critical issue of the approach is the need to report validation results, which can only be based on recorded real life DDoS attacks. There are some public available datasets to perform and report these results. How- ever, results reported on these datasets may not be informative of the system performance on new attacks which may have quite dif- ferent features. This is a pervasive concern in all security applica- tions of machine learning algorithms.

* + 1. *Malware*

Malicious code, such as trojans, virus, spyware, detection by anti-virus approaches can only be performed after some instance of the code has been analyzed ﬁnding some kind of signature, therefore some degree of damage has already been done. Predic- tive approaches based on Machine Learning techniques may al- low anticipative detection at the cost of some false positives. Classiﬁers learn patterns in the known malicious codes extrapo- lating to yet unseen codes. A taxonomy of such approaches is gi- ven in [[183]](#_bookmark102). describing the basic code representation by byte and opcode *n*-grams, strings, and others like portable executable features. Feature selection processes, such as the Fisher score, are applied to ﬁnd the most informative features. Finally, classi- ﬁers tested in this problem include a wide variety of MCS com- bining diverse base classiﬁers with all standard fuser designs. Results have been reported that MCS overcome other ap- proaches, are better suitable for active learning needed to keep the classiﬁers updated and tuned to the changing malicious code versions.

* + 1. *Intrusion detection*

Intrusion Detection and Intrusion Prevention deal with the identiﬁcation of intruder code in a networked environment via the monitoring of communication patterns. Intruder detection per- formed as an anomaly detection process allows to detect previ- ously unseen patterns, at the cost of false alarms, contrary to signature based approaches. The problem is attacked by modular MCS whose compounding base classiﬁers are one-class classiﬁers built by the Parzen window probability density estimation ap- proach [[128]](#_bookmark76). Each module is specialized in a speciﬁc protocol or network service, so that different thresholds can be tuned for each module allowing some optimization of the false alarm rate. On the other hand, Intrusion Prevention tries to impede the execution of the intruder code by fail-safe semantics, automatic response and adaptive enforcement. An approach relies on the fact that Instruc- tion Set Randomization prevents code injection attacks, so that de- tected injected code can be used for adaptation of the anomaly classiﬁer and the signature-based ﬁltering [[184]](#_bookmark104). Clustering of *n*- grams is performed to obtain a model of the normal communica- tion behavior which is accurate allowing zero-day detection of worm infection even in the case of low payload or slow penetration [[185]](#_bookmark106). The interesting proposed hybrid intrusion detection was presented in [[186]](#_bookmark108), where decision trees and support vector ma- chines are combined as a hierarchical hybrid intelligent system model.

* + 1. *Wireless sensor networks*

Wireless sensor networks (WSNs) are collections of inexpen- sive, low power devices deployed over a geographical space for monitoring, measuring and event detection. Anomalies in the WSN can be due to failures in software or hardware, or to mali- cious attacks compelling the sensors to bias or drop their informa- tion and measurements. Anomaly detection in WSN is performed using an ensemble of binary classiﬁers, each tuned on diverse parameters and built following a different approach (Average, autorregresive, neural network, ANFIS). The decision is made by a weighted combination of the classiﬁers outputs [[187]](#_bookmark110).

* 1. *Banking, credit risk, fraud detection*

In the current economical situation, the intelligent processing of ﬁnancial information, the assessing of ﬁnancial or credit risks, and related issues have become a prime concern for society and for the computational intelligence community. Developing new tools may allow to avoid in the future the dire problems faced today by soci- ety. In this section we review some of the most important issues, gathering current attempts to the deal with them.

* + 1. *Fraud detection*

Fraud detection involves identifying fraud as soon as possible after it has been perpetrated. Fraud detection [[188]](#_bookmark111) is big area of research and applications of machine learning, which has provided techniques to counteract fraudsters in credit card fraud, money laundering, telecommunications fraud, and computer intrusion. MCS have been also applied successfully in this domain. A key task is modeling the normal behavior in order to be able to establish suspicion scores for outliers. Probabilistic networks are speciﬁc one-class classiﬁers that are well suited to this task, and bagging of probabilistic networks has been proposed as a general tool for fraud detection because the MCS approach improves the robust- ness of the normal behavior modeling [[189]](#_bookmark112).

* + 1. *Credit card fraud*

Speciﬁc works on credit card fraud detection use real-life data of transactions from an international creditcard operation [[190]](#_bookmark114). The exploration of the sensitivity to the ratio of fraud to non-fraud of the random undersampling approach to deal with unbalanced class sizes is required to validate the approaches. Comparing RF against SVM and logisti regression [[190]](#_bookmark114), RF was the best per- former in all experimental conditions as measured by almost all performance measurements. Other approaches to this problem in- clude a bagged ensemble of SVM tested on a british card applica- tion approval dataset [[191]](#_bookmark115).

* + 1. *Stock market*

Trade based stock market manipulation try to inﬂuence the stock values simply by buying and then selling. It is difﬁcult to de- tect because rules for detection quickly become outdated. An inno- vative research track is the use of peer-group analysis for trade stock manipulation detection, based on the detection of outliers whose dynamic behavior separates from that of the previously similar stock values, its peers [[192]](#_bookmark117). Dynamic clustering allows to track in time the evolution of the community of peers related to the stocks under observation, and outlier detection techniques are required to detect the manipulation events.

* + 1. *Credit risk*

Credit risk prediction models seek to predict whether an indi- vidual will default on a loan or not. It is greatly affected by the unavailability, scarcity and incompleteness of data. The application of machine learning to this problem includes the evaluation of bag- ging, boosting, stacking as well as other conventional classiﬁers

12 *M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17*

over three benchmarking datasets, including sensitivity to noise added to the attributes [[193]](#_bookmark118). Another approach for this problem is the Error Trimmed Boosting (ETB) [[194]](#_bookmark119) which has been tested over a privative dataset provided by a company. ETB consists in the iterative selection of subsets of samples based on their error under the current classiﬁer. An special case of credit risk is enter- prise risk assessment which has a strong economic effect due to the ﬁnancial magnitude of the entities involved. To deal with this problem a combination of bagging and random subspace feature selection using SVM as the base classiﬁer has been developed and tested. The resulting method has increased diversity improv- ing results over a dataset provided by the Bank of China [[195]](#_bookmark121). Bankruptcy prediction is a dramatic special case of credit risk. Ensemble systems with diversity ensured by genetic algorithm based selection of component classiﬁers is proposed in [[196]](#_bookmark122) for bankruptcy prediction in South Korean ﬁrms. The prediction of fail- ure of dotcom companies has been a matter of research since the bubble explosion after the year 2000. Tuning a hybrid of PNN, MLP and genetic programming classiﬁers over a set of features se- lected applying a *t*-test and *F*-test for relevance to the categorical variable has given some solutions [[197]](#_bookmark124). The same approach is re- ported in [[198]](#_bookmark126) to detect fraud in the ﬁnancial statement of big companies.

* + 1. *Financial risks*

Uncertainty in the ﬁnancial operations is identiﬁed with the ﬁnancial risks such as credit, business, investment, and operational risks. Financial distress can be detected by clustering and MCS in four different combination models. Clustering is performed by classical SOM and *k*-means algorithms and used to partition the data space prior to MCS training [[199]](#_bookmark127). Experimental framework for the evaluation of ﬁnancial risk assessment models, giving a spe- ciﬁc performance measures allow the exploration of computational solutions to these problems [[200]](#_bookmark127). Several conventional classiﬁers and MCS have been tested in this framework using a large pool of datasets. Bank performance and bankruptcy prediction is ad- dressed using a widely heterogenous MCS including PNN,RBF,

MLP, SVM, CART trees, and a fuzzy rule system. The effect of PCA

initial dimensionality reduction is also tested [[201]](#_bookmark127). The effect of feature construction from previous experience and *a priori* infor- mation in the efﬁciency of classiﬁers for early warning of bank fail- ures is reported in [[202]](#_bookmark127).

* + 1. *New fraud trends*

Prescription fraud has been identiﬁed as a cause of substantial monetary loss in health care systems, it consists in the prescription of unnecessary medicaments. The research works need to real life data from a large multi-center medical prescription database [[203]](#_bookmark127). The authors use a novel distance based on data-mining approach in a system which is capable of self-learning by regular updates. The system is designed to perform on-line risky prescription detection followed by off-line expert evaluation.

A new brand of frauds appear in the online gaming and lotter- ies, i.e. intended for money laundering, whose detection is dealt with a mixture of supervised and unsupervised classiﬁers [[204]](#_bookmark127). To be adaptive to fraudster evolving strategies, it is required to emphasize online learning, and online cluster detection. Fraud in telecommunication systems involving usage beyond contract spec- iﬁcations is dealt with in [[205]](#_bookmark127) by a preprocessing, clustering and classiﬁcation pipeline. Clustering has been found to improve clas- siﬁcation performance, and boosted trees are the best performing approach. The analysis of social networks by means of MCS may al- low the detection of fraud in automobile insurance, consisting in staging trafﬁc accidents and issuing fake insurance claims to their general or vehicle insurance company [[206]](#_bookmark127).

* 1. *Medicine*

Medicine is a big area of application of any innovative compu- tational approach, dealing with massive amounts of data in some instances, and with very imprecise or ambiguous data in other sit- uations. The range of applications is quite big, so here we only give a scrap of all the current problems and approaches related with the MCS paradigm. In Medicine, a speciﬁc research area since the inception of Artiﬁcial Intelligence is the construction of Computer Aided Diagnosis (CAD) systems or Clinical Decision Support Sys- tems (CDSS) [[207]](#_bookmark127), which involve as the ﬁnal step some kind of classiﬁer predicting the subject’s disease or normal status. In CDSS development, there are several steps such as the deﬁnition of the sensor providing the data, the preprocessing of the data to normal- ize it and remove noise, the selection of features, and the ﬁnal selection of the classiﬁer.

* + 1. *Coronary diseases*

A recent instance of CDSS is the application to cardiovascular disease diagnosis of an heterogenous collection of classiﬁers, com- posed of SVM, bayesian networks and ANN [[208]](#_bookmark128) ﬁnding ten new biomarkers. In this AptaCDSS-E process starts with the use of an aptamer biochip scanning protein expression levels which is the input to physician taking the decisions afterwards. Feature selec- tion is performed by an ANOVA analysis. Doctor decisions are stored for system retraining. Classiﬁer combination is done by majority voting or hierarchical fusion. Many CAD systems related with coronary diseases are based on the information provided by the electrocardiogram (ECG), so that many of them rely on the fea- tures extracted from them. Coronary artery disease is a broad term that encompasses any condition that affects the heart. It is a chronic disease in which the coronary arteries gradually harden and narrow, there have approaches to provide CAD for this condi- tion, such as the use of a mixture of three ANNs for the prediction of coronary artery disease [[209]](#_bookmark129). The dysfunction or abnormality of one or more of the heart four valves is called valvular heart disease. Its diagnosis is performed by neural network ensembles in [[209,210]](#_bookmark129) over features selected by a correlation analysis with the categorical variable. Two separate ANNs are trained to identify myocardial infarction on training sets with different statistics regarding the percentage of patients in [[211]](#_bookmark133). The network special- ized in healthy controls is applied to the new data, if the output is below a threshold the subject is deemed healthy, otherwise the disease-speciﬁc network is applied to decide.

* + 1. *Proteomics*

Proteins are said to have a common fold if they have the same major secondary structure in the same arrangement and with the same topology. Machine learning techniques have been proposed for three-dimensional protein structure prediction. Early ap- proaches consisted in hybrid systems, such as the ANN, statistical classiﬁer and case base reasoning classiﬁer combined by majority voting of [[212]](#_bookmark135). For instance, an ensemble of K-local hyperplanes based on random subspace and feature selection has been tested [[213]](#_bookmark137), where feature selection is done according to distance to the class centroids. A recent approach is the MarFold [[214]](#_bookmark139) com- bining by majority voting three margin-based classiﬁers for protein fold recognition: the adaptive local hyperplane (ALH), the k-neigh- borhood ALH and the SVM.

* + 1. *Neuroscience*

In the ﬁeld of Neurosciences, the machine learning approach is gaining widespread acceptation. It is used for the classiﬁcation of image data searching for predictive non-invasive biomarkers that may allow early or prodromal diagnosis of a number of degenera- tive diseases which have increasing impact in the society due to

*M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17* 13

the aging of populations around the world. Diverse MCS ap- proaches have been applied to structural MRI data, speciﬁcally for the classiﬁcation of Alzheimer disease patients, such as an RVM based two stage pipeline [[45]](#_bookmark50), variations of Adaboost [[215]](#_bookmark130), hybridizations of kernel and Dendritic Computing approaches [[216]](#_bookmark130). Classiﬁer Ensembles have been applied to the classiﬁcation of fMRI data [[217,218]](#_bookmark130) and its visual decoding [[219]](#_bookmark130), which is the reconstruction of the visual stimuli from the fMRI data.

* 1. *Recommender systems*

Nowadays, recommender systems are the focus of intense re- search [[220]](#_bookmark130). They try to help consumers to select the product that may be interesting for them based on their previous searches and transactions, but such systems are expanding beyond typical sales. They are used to predict which mobile telephone subscribers are in risk of switching to another provider, or to advice conference orga- nizers about assigning papers to peer reviewers [[221]](#_bookmark130). Burke [[222]](#_bookmark130) proposed hybrid recommender systems combining two or more recommendation techniques to improve performance avoiding the drawbacks of an individual recommender. Similar observations were conﬁrmed by Balabanovic et al. [[223]](#_bookmark130) and Pazzani [[224]](#_bookmark130) who demonstrated that hybrid method recommentations improve col- laborative and content-based approaches.

There are several interesting works which apply the hybrid and

combined approach to recommender systems. Jahrer and Töscher

[[225]](#_bookmark130) demonstrated the advantage of ensemble learning applied to the combination of different collaborative ﬁltering algorithms on the Netix Prize dataset. Porcel et al. [[226]](#_bookmark131) developed an hybrid fuzzy recommender system to help disseminate information about research resources in the ﬁeld of interest of a user. Claypool et al.

[[227]](#_bookmark132) performed a linear combination of the ratings obtained from individual recommender systems into one ﬁnal recommendation, while Pazzani proposed to use a voting scheme [[224]](#_bookmark130). Billsus and Pazzani [[228]](#_bookmark134) selected the best recommendation on the basis of a recommendation quality metric as the level of conﬁdence while Tran and Cohen [[229]](#_bookmark136) preferred an individual which is the most consistent with the previous ratings of the user. Kunaver et al.

[[230]](#_bookmark138) proposed Combined Collaborative Recommender based on three different collaborative recommender techniques. Goksedef and Gundoz-Oguducu [[231]](#_bookmark140) combined the results of several rec- ommender techniques based on Web usage mining.

1. Final remarks

We have summarized the main research streams on multiple classiﬁer systems, also known in the literature as combined classi- ﬁer or classiﬁer ensemble. Such hybrid systems are the focus of in- tense research recently, so fruitful that our review could not be exhaustive. Key issues related to the problem under consideration are classiﬁer diversity and methods of classiﬁer combination.

The diversity is believed to provide improved accuracy and clas- siﬁer performance. Most works try to obtain maximum diversity by different means: introducing classiﬁer heterogeneity, boot- strapping the training data, randomizing feature selection, ran- domizing subspace projections, boosting the data weights, and many combinations of these ideas. Nowadays, the diversity hypothesis has not been fully proven, either theoretically or empir- ically. However, the fact is that MCSs show in most instances im- proved performance, resilience and robustness to high data dimensionality and diverse forms of noise, such as labeling noise. The there are several propositions how to combine the classiﬁer outputs, what was presented in this work, nonetheless we point out that classiﬁer combination is not the only way to produce hy-

brid classiﬁer systems. We envisage further possibilities of hybrid- ization such as:

Merging the raw data from different sources into one repository and then train the classiﬁer.

●

Merging the raw data and a *prior* expert knowledge (e.g., learn- ing sets and human expert rules to improve rules on the basis of incoming data).

●

Merging a prior expert knowledge and classiﬁcation models returned by machine learning procedures.

●

For such a problem we have to take into consideration issues re- lated to data privacy, computational and memory efﬁciency.

Acknowledgments

We would like to thank the anonymous reviewers for their dil- igent work and efﬁcient efforts. We are also grateful to the Editor- in-Chief, Prof. Belur V. Dasarathy, who encouraged us to write this survey for this prestigious journal.

MichałWoz´niak was supported by The Polish National Science Centre under the Grant No. N519 576638 which is being realized in years 2010–2013.

References

1. [J. Neumann, The Computer and the Brain,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0005) [Yale University Press, New Haven,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0005) [CT, USA, 1958](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0005).
2. [A. Newell, Intellectual issues in the history of artiﬁcial intelligence, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0010) [F.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0010) [Machlup,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0010) [U. Mansﬁeld (Eds.), The Study of Information: Interdisciplinary](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0010) [Messages,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0010) [John Wiley & Sons Inc., New York, NY, USA, 1983, pp. 187–294](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0010).
3. D. Wolpert, The supervised learning no-free-lunch theorems, in: Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications, 2001, pp. 25–42.
4. [C.K. Chow, Statistical independence and threshold functions, IEEE](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0015) [Transactions on Electronic Computers EC-14 (1) (1965) 66–68](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0015).
5. [L. Shapley, B. Grofman, Optimizing group judgmental accuracy in the](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0020) [presence of interdependencies, Public Choice 43 (3) (1984) 329–333](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0020).
6. [B.V. Dasarathy, B.V. Sheela, A composite classiﬁer system design: concepts](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0025) [and methodology, Proceedings of the IEEE 67 (5) (1979) 708–713](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0025).
7. L. Rastrigin, R.H. Erenstein, Method of Collective Recognition, Energoizdat, Moscow, 1981.
8. L. Hansen, P. Salamon, Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (10) (1990) 993–1001, http:// [dx.doi.org/10.1109/34.58871](http://dx.doi.org/10.1109/34.58871).
9. [L. Xu, A. Krzyzak, C. Suen, Methods of combining multiple classiﬁers and their](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0035) [applications to handwriting recognition, IEEE Transactions on Systems, Man](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0035) [and Cybernetics 22 (3) (1992) 418–435](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0035).
10. [K. Tumer, J. Ghosh, Analysis of decision boundaries in linearly combined](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0040) [neural classiﬁers, Pattern Recognition 29 (2) (1996) 341–348](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0040).
11. [T. Ho, J.J. Hull, S. Srihari, Decision combination in multiple classiﬁer systems,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0045) [IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (1) (1994)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0045) [66–75](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0045).
12. [L. Breiman, Bagging predictors, Machine Learning 24 (2) (1996) 123–140](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0050).
13. [R. Schapire, The strength of weak learnability, Machine Learning 5 (2) (1990)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0055) [197–227](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0055).
14. [Y. Freund, Boosting a weak learning algorithm by majority, Information](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0060) [Computing 121 (2) (1995) 256–285](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0060).
15. [M. Kearns, U. Vazirani, An Introduction to Computational Learning Theory,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0065) [MIT Press, Cambridge, MA, USA, 1994](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0065).
16. [D. Angluin, Queries and concept learning, Machine Learning 2 (4) (1988) 319–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0070) [342](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0070).
17. [A. Jain, R. Duin, M. Jianchang, Statistical pattern recognition: a review, IEEE](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0075) [Transactions on Pattern Analysis and Machine Intelligence 22 (1) (2000) 4–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0075) [37](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0075).
18. [N. Oza, K. Tumer, Classiﬁer ensembles: select real-world applications,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0080) [Information Fusion 9 (1) (2008) 4–20](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0080).
19. [R. Polikar, Ensemble based systems in decision making, IEEE Circuits and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0085) [Systems Magazine 6 (3) (2006) 21–45](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0085).
20. [R. Polikar, Ensemble learning, Scholarpedia 3 (12) (2008) 2776](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0090).
21. [L. Rokach, Taxonomy for characterizing ensemble methods in classiﬁcation](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0095) [tasks: a review and annotated bibliography, Computational Statistics and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0095) [Data Analysis 53 (12) (2009) 4046–4072](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0095).
22. [L. Kuncheva, Combining Pattern Classiﬁers: Methods and Algorithms, Wiley-](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0100) [Interscience, 2004](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0100).
23. [L. Rokach, Pattern Classiﬁcation Using Ensemble Methods, Series in Machine](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0105) [Perception and Artiﬁcial Intelligence,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0105) [World Scientiﬁc, 2010](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0105).

14 *M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17*

1. [G. Seni, J. Elder, Ensemble Methods in Data Mining: Improving Accuracy](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0110) [Through Combining Predictions,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0110) [Morgan and Claypool Publishers, 2010](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0110).
2. [B. Baruque, E. Corchado, Fusion Methods for Unsupervised Learning](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0115) [Ensembles, Springer Verlag New York, Inc., 2011](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0115).
3. [R. Duda, P. Hart, D. Stork, Pattern Classiﬁcation, second ed., Wiley, New York,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0120) [2001](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0120).
4. [E. Alpaydin, Introduction to Machine Learning, second ed., The MIT Press,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0125) [2010](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0125).
5. [C. Bishop, Pattern Recognition and Machine Learning (Information Science](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0130) [and Statistics),](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0130) [Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0130).
6. [T. Dietterich, Ensemble methods in machine learning, in: Multiple Classiﬁer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0135) [Systems, Lecture Notes in Computer Science, vol. 1857, Springer, Berlin,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0135) [Heidelberg, 2000, pp. 1–15](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0135).
7. G. Marcialis, F. Roli, Fusion of face recognition algorithms for video-based surveillance systems, in: G.L. Foresti, C. Regazzoni, P. Varshney (Eds.), 2003, pp. 235–250.
8. [S. Hashem, Optimal linear combinations of neural networks, Neural Networks](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0140) [10 (4) (1997) 599–614](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0140).
9. [R. Clemen, Combining forecasts: a review and annotated bibliography,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0145) [International Journal of Forecasting 5 (4) (1989) 559–583](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0145).
10. [J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Series in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0150) [Machine Learning,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0150) [Morgan Kaufman Publishers, 1993](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0150).
11. [T. Wilk, M. Wozniak, Complexity and multithreaded implementation analysis](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0155) [of one class-classiﬁers fuzzy combiner, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0155) [E. Corchado,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0155) [M. Kurzynski,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0155) [M.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0155) [Wozniak (Eds.), Hybrid Artiﬁcial Intelligent Systems, Lecture Notes in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0155) [Computer Science, vol. 6679, Springer, Berlin/Heidelberg, 2011, pp. 237–244](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0155).
12. T. Kacprzak, K. Walkowiak, M. Wozniak, Optimization of overlay distributed computing systems for multiple classiﬁer system – heuristic approach, Logic Journal of IGPL, doi:10.1093/jigpal/jzr020.
13. [K. Walkowiak, Anycasting in connection-oriented computer networks:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0160) [models, algorithms and results, International Journal of Applied](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0160) [Mathematics and Computer Sciences 20 (1) (2010) 207–220](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0160).
14. [R. Agrawal, R. Srikant, Privacy-preserving data mining, SIGMOD Records 29](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0165) [(2) (2000) 439–450](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0165).
15. G. Giacinto, F. Roli, G. Fumera, Design of effective multiple classiﬁer systems by clustering of classiﬁers, in: Proceedings of the 15th International Conference on Pattern Recognition, 2000, vol. 2, 2000, pp. 160–163.
16. [T. Ho, Complexity of classiﬁcation problems and comparative advantages of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0170) [combined classiﬁers, in: Proceedings of the First International Workshop on](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0170) [Multiple Classiﬁer Systems, MCS ’00,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0170) [Springer-Verlag, London, UK, 2000, pp.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0170) [97–106](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0170).
17. [F. Roli, G. Giacinto, Design of Multiple Classiﬁer Systems,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0175) [World Scientiﬁc](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0175) [Publishing, 2002](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0175).
18. [L. Lam, Classiﬁer combinations: implementations and theoretical issues, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0180) [Proceedings of the First International Workshop on Multiple Classiﬁer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0180) [Systems, MCS ’00,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0180) [Springer-Verlag, London, UK, 2000, pp. 77–86](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0180).
19. [A.F.R. Rahman, M.C. Fairhurst, Serial combination of multiple experts: a](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0185) [uniﬁed evaluation, Pattern Analysis and Applications 2 (1999) 292–311](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0185).
20. [G. Fumera, I. Pillai, F. Roli, A two-stage classiﬁer with reject option for text](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0190) [categorisation, 5th International Workshop on Statistical Techniques in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0190) [Pattern Recognition (SPR 2004), vol. 3138, Springer, Lisbon, Portugal, 2004,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0190) [pp. 771–779](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0190).
21. [P. Bartlett, M. Wegkamp, Classiﬁcation with a reject option using a hinge loss,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0195) [Journal of Machine Learning Research 9 (2008) 1823–1840](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0195).
22. [M. Termenon, M. Graña, A two stage sequential ensemble applied to the](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0200) [classiﬁcation of alzheimer’s disease based on MRI features, Neural Processing](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0200) [Letters 35 (1) (2012) 1–12](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0200).
23. [P. Clark, T. Niblett, The CN2 induction algorithm, Machine Learning 3 (4)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0205) [(1989) 261–283](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0205).
24. [R. Rivest, Learning decision lists, Machine Learning 2 (3) (1987) 229–246](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0210).
25. Y. Freund, R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences 55 (1) (1997) 119–139, <http://dx.doi.org/10.1006/jcss.1997.1504>.
26. X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.-H. Zhou, M. Steinbach, D.J. Hand, D. Steinberg, Top 10 algorithms in data mining, Knowledge and Information Systems 14 (1) (2008) 1–37, <http://dx.doi.org/10.1007/s10115-007-0114-2>.
27. R. Schapire, The strength of weak learnability, Machine Learning 5 (2) (1990) 197–227, http://dx.doi.org/10.1023/A:1022648800760.
28. J. Kivinen, M.K. Warmuth, Boosting as entropy projection, in: Proceedings of the Twelfth Annual Conference on Computational Learning Theory, 1999.

<<http://dl.acm.org/citation.cfm?id=307424>>.

1. [D. Partridge, W. Krzanowski, Software diversity: practical statistics for its](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0230) [measurement and exploitation, Information and Software Technology 39 (10)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0230) [(1997) 707–717](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0230).
2. G. Brown, L. Kuncheva, ‘‘good’’ And ‘‘bad’’ diversity in majority vote ensembles, in: Proceedings MCS 2010, pp. 124–133.
3. [M. Smetek, B. Trawinski, Selection of heterogeneous fuzzy model ensembles](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0235) [using self-adaptive genetic algorithms, New Generation Computing 29 (2011)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0235) [309–327](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0235).
4. [A.J.C. Sharkey, N. Sharkey, Combining diverse neural nets, Knowledge](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0240) [Engineering Review 12 (3) (1997) 231–247](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0240).
5. [A. Krogh, J. Vedelsby, Neural network ensembles, cross validation, and active](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0245) [learning, Advances in Neural Information Processing Systems 7 (1995) 231–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0245) [238](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0245).
6. [G. Zenobi, P. Cunningham, Using diversity in preparing ensembles of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0250) [classiﬁers based on different feature subsets to minimize generalization](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0250) [error, Machine Learning: ECML 2001 (2001) 576–587](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0250).
7. [G. Brown, J. Wyatt, R. Harris, X. Yao, Diversity creation methods: a survey and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0255) [categorisation, Information Fusion 6 (1) (2005) 5–20](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0255).
8. N. Ueda, R. Nakano, Generalization error of ensemble estimators, in: Proceedings of IEEE International Conference on Neural Networks, Washington, USA, 1996, pp. 90–95.
9. [G. Brown, J. Wyatt, P. Tinˇo, Managing diversity in regression ensembles,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0260) [Journal of Machine Learning Research 6 (2005) 1621–1650](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0260).
10. [L. Kuncheva, C. Whitaker, C. Shipp, R. Duin, Limits on the majority vote](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0265) [accuracy in classiﬁer fusion, Pattern Analysis and Applications 6 (2003) 22–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0265) [31](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0265).
11. [Y. Bi, The impact of diversity on the accuracy of evidential classiﬁer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0270) [ensembles, International Journal of Approximate Reasoning 53 (4) (2012)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0270) [584–607](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0270).
12. [D. Margineantu, T. Dietterich, Pruning adaptive boosting, in: Proceedings of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0275) [the Fourteenth International Conference on Machine Learning, ICML ’97,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0275) [Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997, pp. 211–218](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0275).
13. D. Skalak, The sources of increased accuracy for two proposed boosting algorithms, in: Proceedings of the American Association for Arti Intelligence, AAAI-96, Integrating Multiple Learned Models Workshop, 1996, pp. 120–125.
14. [G. Giacinto, F. Roli, Design of effective neural network ensembles for image](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0280) [classiﬁcation purposes, Image Vision Computing 19 (9-10) (2001) 699–707](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0280).
15. R. Kohavi, D. Wolpert, Bias plus variance decomposition for zero-one loss functions, in: ICML-96, 1996.
16. [J. Fleiss, J. Cuzick, The reliability of dichotomous judgments: unequal](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0285) [numbers of judgments per subject, Applied Psychological Measurement 4](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0285) [(3) (1979) 537–542](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0285).
17. [P. Cunningham, J. Carney, Diversity versus quality in classiﬁcation ensembles](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0290) [based on feature selection, in: Proceedings of the 11th European Conference](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0290) [on Machine Learning, ECML ’00, Springer-Verlag, London, UK, 2000, pp. 109–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0290) [116](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0290).
18. [C. Shipp, L. Kuncheva, Relationships between combination methods and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0295) [measures of diversity in combining classiﬁers, Information Fusion 3 (2)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0295) [(2002) 135–148](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0295).
19. [E.K. Tang, P.N. Suganthan, X. Yao, An analysis of diversity measures, Machine](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0300) [Learning 65 (1) (2006) 247–271](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0300).
20. [G. Martinez-Mu/ noz, D. Hern/’andez-Lobato, A. Suarez, An analysis of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0305) [ensemble pruning techniques based on ordered aggregation, IEEE](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0305) [Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0305) [245–259](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0305).
21. [D. Ruta, B. Gabrys, Classiﬁer selection for majority voting, Information Fusion](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0310) [6 (1) (2005) 63–81](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0310).
22. [R. Banﬁeld, L. Hall, K. Bowyer, W. Kegelmeyer, Ensemble diversity measures](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0315) [and their application to thinning, Information Fusion 6 (1) (2005) 49–62](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0315).
23. [Z.-H. Zhou, J. Wu, W. Tang, Ensembling neural networks: many could be](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0320) [better than all, Artiﬁcial Intelligence 137 (1-2) (2002) 239–263](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0320).
24. [B. Gabrys, D. Ruta, Genetic algorithms in classiﬁer fusion, Applied Soft](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0325) [Computing 6 (4) (2006) 337–347](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0325).
25. [I. Partalas, G. Tsoumakas, I. Vlahavas, Pruning an ensemble of classiﬁers via](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0330) [reinforcement learning, Neurocomputing 72 (7–9) (2009) 1900–1909](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0330).
26. Q. Dai, A competitive ensemble pruning approach based on cross-validation technique, Knowledge-Based Systems (0) (2012), [http://dx.doi.org/10.1016/](http://dx.doi.org/10.1016/j.knosys.2012.08.024) [j.knosys.2012.08.024](http://dx.doi.org/10.1016/j.knosys.2012.08.024).
27. [Y. Peng, Q. Huang, P. Jiang, J. Jiang, Cost-sensitive ensemble of support vector](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340) [machines for effective detection of microcalciﬁcation in breast cancer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340) [diagnosis, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340) [L. Wang,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340) [Y. Jin (Eds.), Fuzzy Systems and Knowledge](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340) [Discovery, Lecture Notes in Computer Science, vol. 3614,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340) [Springer, Berlin/](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340) [Heidelberg, 2005, pp. 483–493](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0340).
28. [K. Jackowski, B. Krawczyk, M. Woniak, Cost-sensitive splitting and selection](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0345) [method for medical decision support system, in: H. Yin, J.A. Costa, G. Barreto](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0345) [(Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2012,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0345) [Lecture Notes in Computer Science, vol. 7435,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0345) [Springer, Berlin Heidelberg,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0345) [2012, pp. 850–857](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0345).
29. W. Du, Z. Zhan, Building decision tree classiﬁer on private data, in: Proceedings of the IEEE International Conference on Privacy, Security and Data Mining – Volume 14, CRPIT ’14, Australian Computer Society, Inc., Darlinghurst, Australia, 2002, pp. 1–8.
30. [B. Krawczyk, M. Wozniak, Privacy preserving models of k-NN algorithm, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350)

[R. Burduk,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350) [M. Kurzynski,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350) [M. Wozniak,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350) [A. Zolnierek (Eds.), Computer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350) [Recognition Systems 4, Advances in Intelligent and Soft Computing, vol.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350) [95,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350) [Springer, Berlin/Heidelberg, 2011, pp. 207–217](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0350).

1. [Y. Lindell, B. Pinkas, Secure multiparty computation for privacy-preserving](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0355) [data mining, IACR Cryptology ePrint Archive 2008 (2008) 197](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0355).
2. [K. Walkowiak, S. Sztajer, M. Wozniak, Decentralized distributed computing](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [system for privacy-preserving combined classiﬁers – modeling and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [optimization, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [B. Murgante,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [O. Gervasi,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [A. Iglesias,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [D. Taniar,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [B. Apduhan](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [(Eds.), Computational Science and Its Applications – ICCSA 2011, Lecture](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [Notes in Computer Science, Vol. 6782,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [Springer, Berlin/Heidelberg, 2011, pp.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360) [512–525](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0360).
3. [A. Pavlo, E. Paulson, A. Rasin, D. Abadi, D. DeWitt, S. Madden, M. Stonebraker,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0365) [A comparison of approaches to large-scale data analysis, in: Proceedings of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0365) [the 2009 ACM SIGMOD International Conference on Management of Data,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0365) [SIGMOD ’09, ACM, New York, NY, USA, 2009, pp. 165–178](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0365).

*M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17* 15

1. R.E. Schapire, The boosting approach to machine learning: an overview, in: MSRI Workshop on Nonlinear Estimation and Classiﬁcation, Berkeley, CA, USA, 2001.
2. T. Ho, Random decision forests, in: Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)–Volume 1, ICDAR ’95, IEEE Computer Society, Washington, DC, USA, 1995, pp. 278–.
3. [T. Ho, The random subspace method for constructing decision forests, IEEE](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0370) [Transactions on Pattern Analysis and Machine Intelligence 20 (1998) 832–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0370) [844](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0370).
4. [L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0375).
5. [M. Skurichina, R. Duin, Bagging, boosting and the random subspace method](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0380) [for linear classiﬁers, Pattern Analysis and Applications 5 (2) (2002) 121–135](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0380).
6. G. Tremblay, R. Sabourin, P. Maupin, Optimizing nearest neighbour in random subspaces using a multi-objective genetic algorithm, in: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR’04) Volume 1– Volume 01, ICPR ’04, IEEE Computer Society, Washington, DC, USA, 2004, pp. 208–.
7. [S. Bay, Nearest neighbor classiﬁcation from multiple feature subsets,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0385) [Intelligent Data Analysis 3 (3) (1999) 191–209](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0385).
8. [L. Nanni, Letters: Experimental comparison of one-class classiﬁers for online](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0390) [signature veriﬁcation, Neurocomputing 69 (7–9) (2006) 869–873](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0390).
9. [D. Tao, X. Tang, X. Li, X. Wu, Asymmetric bagging and random subspace for](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0395) [support vector machines-based relevance feedback in image retrieval, IEEE](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0395) [Transactions on Pattern Analysis Machine Intelligence 28 (7) (2006) 1088–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0395) [1099](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0395).
10. [K. Ting, J. Wells, S. Tan, S. Teng, G. Webb, Feature-subspace aggregating:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0400) [ensembles for stable and unstable learners, Machine Learning 82 (2011) 375–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0400) [397](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0400).
11. [R. Bryll, R. Gutierrez-Osuna, F. Quek, Attribute bagging: improving accuracy](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0405) [of classiﬁer ensembles by using random feature subsets, Pattern Recognition](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0405) [36 (6) (2003) 1291–1302](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0405).
12. [Y. Baram, Partial classiﬁcation: the beneﬁt of deferred decision, IEEE](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0410) [Transactions on Pattern Analysis and Machine Intelligence 20 (8) (1998)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0410) [769–776](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0410).
13. [L. Cordella, P. Foggia, C. Sansone, F. Tortorella, M. Vento, A cascaded multiple](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0415) [expert system for veriﬁcation, in: Multiple Classiﬁer Systems, Lecture Notes](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0415) [in Computer Science, vol. 1857, Springer, Berlin/Heidelberg, 2000, pp. 330–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0415) [339](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0415).
14. K. Goebel, W. Yan, Choosing classiﬁers for decision fusion, in: Proceedings of the Seventh International Conference on Information Fusion, 2004, pp. 563– 568.
15. [B. Baruque, S. Porras, E. Corchado, Hybrid classiﬁcation ensemble using](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0420) [topology-preserving clustering, New Generation Computing 29 (2011) 329–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0420) [344](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0420).
16. L. Kuncheva, Clustering-and-selection model for classiﬁer combination, in: Proceedings of the Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000, vol. 1, 2000, pp. 185–188.
17. [K. Jackowski, M. Wozniak, Algorithm of designing compound recognition](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0425) [system on the basis of combining classiﬁers with simultaneous splitting](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0425) [feature space into competence areas, Pattern Analysis and Applications 12 (4)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0425) [(2009) 415–425](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0425).
18. M. Wozniak, B. Krawczyk, Combined classiﬁer based on feature space partitioning, International Journal of Applied Mathematics and Computer Sciences 22 (4) (2012) 855–866.
19. [H. Lee, C. Chen, J. Chen, Y. Jou, An efﬁcient fuzzy classiﬁer with feature](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0430) [selection based on fuzzy entropy, IEEE Transactions on Systems, Man, and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0430) [Cybernetics, Part B: Cybernetics 31 (3) (2001) 426–432](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0430).
20. [J. Hong, J. Min, U. Cho, S. Cho, Fingerprint classiﬁcation using one-vs-all](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0435) [support vector machines dynamically ordered with naïve bayes classiﬁers,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0435) [Pattern Recognition 41 (2008) 662–671](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0435).
21. [A.R. Ko, R. Sabourin, A. Britto, From dynamic classiﬁer selection to dynamic](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0440) [ensemble selection, Pattern Recognition 41 (5) (2008) 1735–1748](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0440).
22. [L. Didaci, G. Giacinto, F. Roli, G. Marcialis, A study on the performances of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0445) [dynamic classiﬁer selection based on local accuracy estimation, Pattern](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0445) [Recognition 38 (11) (2005) 2188–2191](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0445).
23. [G. Giacinto, F. Roli, Dynamic classiﬁer selection based on multiple classiﬁer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0450) [behavior, Pattern Recognition 34 (9) (2001) 1879–1881](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0450).
24. M. de Souto, R. Soares, A. Santana, A. Canuto, Empirical comparison of dynamic classiﬁer selection methods based on diversity and accuracy for building ensembles, in: IJCNN 2008, IEEE International Joint Conference on Neural Networks, 2008, IEEE World Congress on Computational Intelligence, 2008, pp. 1480–1487.
25. [T. Woloszynski, M. Kurzynski, A probabilistic model of classiﬁer competence](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0455) [for dynamic ensemble selection, Pattern Recognition 44 (1011) (2011) 2656–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0455) [2668](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0455).
26. [T. Woloszynski, M. Kurzynski, P. Podsiadlo, G. Stachowiak, A measure of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0460) [competence based on random classiﬁcation for dynamic ensemble selection,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0460) [Information Fusion 13 (3) (2012) 207–213](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0460).
27. [W. Street, Y. Kim, A streaming ensemble algorithm (sea) for large-scale](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0465) [classiﬁcation, in: Proceedings of the Seventh ACM SIGKDD International](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0465) [Conference on Knowledge Discovery and Data Mining, KDD ’01,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0465) [ACM, New](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0465) [York, NY, USA, 2001, pp. 377–382](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0465).
28. [H. Wang, W. Fan, P. Yu, J. Han, Mining concept-drifting data streams using](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0470) [ensemble classiﬁers, in: Proceedings of the Ninth ACM SIGKDD International](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0470)

[Conference on Knowledge Discovery and Data Mining, KDD ’03,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0470) [ACM, New](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0470) [York, NY, USA, 2003, pp. 226–235](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0470).

1. [Y. Zhang, X. Jin, An automatic construction and organization strategy for](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0475) [ensemble learning on data streams, SIGMOD Record 35 (3) (2006) 28–33](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0475).
2. J. Kolter, M. Maloof, Dynamic weighted majority: a new ensemble method for tracking concept drift, in: ICDM 2003, Third IEEE International Conference on Data Mining, 2003, 2003, pp. 123–130.
3. [A. Tsymbal, M. Pechenizkiy, P. Cunningham, S. Puuronen, Dynamic](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0480) [integration of classiﬁers for handling concept drift, Information Fusion 9](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0480) [(1) (2008) 56–68](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0480).
4. [X. Zhu, X. Wu, Y. Yang, Effective classiﬁcation of noisy data streams with](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0485) [attribute-oriented dynamic classiﬁer selection, Knowledge Information](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0485) [Systems 9 (3) (2006) 339–363](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0485).
5. D. Tax, R. Duin, Using two-class classiﬁers for multiclass classiﬁcation, in: Proceedings of the 16th International Conference on Pattern Recognition, 2002, vol. 2, 2002, pp. 124 –127.
6. [T. Dietterich, G. Bakiri, Solving multiclass learning problems via error-](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0490) [correcting output codes, Journal of Artiﬁcial Intelligence Research 2 (1995)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0490) [263–286](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0490).
7. [K. Duan, S. Keerthi, W. Chu, S. Shevade, A. Poo, Multi-category classiﬁcation](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0495) [by soft-max combination of binary classiﬁers, in: Proceedings of the 4th](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0495) [International Conference on Multiple Classiﬁer Systems, MCS’03,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0495) [Springer-](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0495) [Verlag, Berlin, Heidelberg, 2003, pp. 125–134](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0495).
8. [A. Passerini, M. Pontil, P. Frasconi, New results on error correcting output](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0500) [codes of kernel machines, IEEE Transactions on Neural Networks 15 (1)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0500) [(2004) 45–54](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0500).
9. [T. Wu, C. Lin, R. Weng, Probability estimates for multi-class classiﬁcation by](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0505) [pairwise coupling, Journal of Machine Learning Research 5 (2004) 975–1005](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0505).
10. J. Friedman, Another Approach to Polychotomous Classiﬁcation, Tech. rep., Department of Statistics, Stanford University, 1996.
11. [E. Hüllermeier, S. Vanderlooy, Combining predictions in pairwise](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0510) [classiﬁcation: an optimal adaptive voting strategy and its relation to](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0510) [weighted voting, Pattern Recognition 43 (1) (2010) 128–142](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0510).
12. [M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, An overview of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0515) [ensemble methods for binary classiﬁers in multi-class problems:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0515) [Experimental study on one-vs-one and one-vs-all schemes, Pattern](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0515) [Recognition 44 (8) (2011) 1761–1776](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0515).
13. D. Tax, R.P.W. Duin, Characterizing one-class datasets, in: Proceedings of the Sixteenth Annual Symposium of the Pattern Recognition Association of South Africa, 2005, pp. 21–26.
14. [D. Tax, R. Duin, Combining one-class classiﬁers, in: Proceedings of the Second](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0520) [International Workshop on Multiple Classiﬁer Systems, MCS ’01,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0520) [Springer-](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0520) [Verlag, London, UK, 2001, pp. 299–308](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0520).
15. [T. Wilk, M. Wozniak, Soft computing methods applied to combination of one-](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0525) [class classiﬁers, Neurocomputing 75 (2012) 185–193](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0525).
16. [G. Giacinto, R. Perdisci, M. Del Rio, F. Roli, Intrusion detection in computer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0530) [networks by a modular ensemble of one-class classiﬁers, Information Fusion](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0530) [9 (2008) 69–82](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0530).
17. [Y. Hu, Handbook of Neural Network Signal Processing, 1st ed., CRC Press, Inc.,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0535) [Boca Raton, FL, USA, 2000](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0535).
18. [K. Woods, W.P. Kegelmeyer Jr., K. Bowyer, Combination of multiple classiﬁers](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0540) [using local accuracy estimates, IEEE Transactions on Pattern Analysis and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0540) [Machine Intelligence 19 (4) (1997) 405–410](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0540).
19. [M. Wozniak, M. Zmyslony, Combining classiﬁers using trained fuser –](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0545) [analytical and experimental results, Neural Network World 13 (7) (2010)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0545) [925–934](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0545).
20. [S. Raudys, Trainable fusion rules. I. Large sample size case, Neural Networks](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0550) [19 (10) (2006) 1506–1516](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0550).
21. M. van Erp, L. Vuurpijl, L. Schomaker, An overview and comparison of voting methods for pattern recognition, in: Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition, 2002, 2002, pp. 195–200.
22. [M. Wozniak, K. Jackowski, Some remarks on chosen methods of classiﬁer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0555) [fusion based on weighted voting, in: E. Corchado, X. Wu, E. Oja, A. Herrero, B.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0555) [Baruque (Eds.), Hybrid Artiﬁcial Intelligence Systems, Lecture Notes in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0555) [Computer Science, vol. 5572, Springer, Berlin/Heidelberg, 2009, pp. 541–548](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0555).
23. [S. Raudys, Trainable fusion rules. II. Small sample-size effects, Neural](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0560) [Networks 19 (10) (2006) 1517–1527](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0560).
24. [H. Inoue, H. Narihisa, Optimizing a multiple classiﬁer system, in: M. Ishizuka,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0565)

[A. Sattar (Eds.), PRICAI 2002: Trends in Artiﬁcial Intelligence, Lecture Notes in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0565) [Computer Science, vol. 2417,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0565) [Springer, Berlin/Heidelberg, 2002, pp. 1–16](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0565).

1. L. Alexandre, A. Campilho, M. Kamel, Combining independent and unbiased classiﬁers using weighted average., in: Proceedings ICPR 2000, 2000, pp. 2495–2498.
2. [B. Biggio, G. Fumera, F. Roli, Bayesian analysis of linear combiners, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0570) [Proceedings of the 7th International Conference on Multiple Classiﬁer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0570) [Systems, MCS ’07,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0570) [Springer-Verlag, Berlin, Heidelberg, 2007, pp. 292–301](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0570).
3. [J. Kittler, F. Alkoot, Sum versus vote fusion in multiple classiﬁer systems, IEEE](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0575) [Transactions on Pattern Analysis and Machine Intelligence 25 (1) (2003)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0575) [110–115](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0575).
4. [N. Rao, A generic sensor fusion problem: classiﬁcation and function](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580) [estimation, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580) [F. Roli,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580) [J. Kittler,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580) [T. Windeatt (Eds.), Multiple Classiﬁer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580) [Systems, Lecture Notes in Computer Science, vol. 3077,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580) [Springer, 2004, pp.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580) [16–30](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0580).
5. D. Opitz, J. Shavlik, Generating accurate and diverse members of a neural- network ensemble, in: NIPS, 1995, pp. 535–541.

16 *M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17*

1. [L. Rokach, O. Maimon, Feature set decomposition for decision trees,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0585) [Intelligent Data Analysis 9 (2) (2005) 131–158](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0585).
2. G. Fumera, F. Roli, A theoretical and experimental analysis of linear combiners for multiple classiﬁer systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (6) (2005) 942–956, <http://dx.doi.org/> [10.1109/TPAMI.2005.109](http://dx.doi.org/10.1109/TPAMI.2005.109).
3. [M. Wozniak, Experiments on linear combiners, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0595) [E. Pietka,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0595) [J. Kawa (Eds.),](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0595) [Information Technologies in Biomedicine, Advances in Soft Computing, vol.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0595) [47,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0595) [Springer, Berlin/Heidelberg, 2008, pp. 445–452](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0595).
4. R. Duin, The combining classiﬁer: to train or not to train? in: Proceedings of the 16th International Conference on Pattern Recognition, 2002, vol. 2, 2002, pp. 765–770.
5. [R. Jacobs, M. Jordan, S. Nowlan, G. Hinton, Adaptive mixtures of local experts,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0600) [Neural Computation 3 (1991) 79–87](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0600).
6. [R. Jacobs, Methods for combining experts’ probability assessments, Neural](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0605) [Computation 7 (5) (1995) 867–888](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0605).
7. [V. Tresp, M. Taniguchi, Combining estimators using non-constant weighting](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0610) [functions, Advances in Neural Information Processing Systems, vol. 7, MIT](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0610) [Press, 1995, pp. 419–426](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0610).
8. [P. Cheeseman, M. Self, J. Kelly, J. Stutz, W. Taylor, D. Freeman, AutoClass: a](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0615) [Bayesian classiﬁcation system, in: Machine Learning: Proceedings of the Fifth](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0615) [International Workshop,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0615) [Morgan Kaufman, 1988](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0615).
9. [S. Shlien, Multiple binary decision tree classiﬁers, Pattern Recognition 23 (7)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0620) [(1990) 757–763](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0620).
10. [M. Wozniak, Experiments with trained and untrained fusers, in: E. Corchado,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0625)

[J. Corchado,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0625) [A. Abraham (Eds.), Innovations in Hybrid Intelligent Systems,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0625) [Advances in Soft Computing, vol. 44,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0625) [Springer, Berlin/Heidelberg, 2007, pp.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0625) [144–150](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0625).

1. [M. Wozniak, Evolutionary approach to produce classiﬁer ensemble based on](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0630) [weighted voting, in: NaBIC 2009, World Congress on Nature & Biologically](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0630) [Inspired Computing, 2009,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0630) [IEEE, 2009, pp. 648–653](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0630).
2. L. Lin, X. Wang, B. Liu, Combining multiple classiﬁers based on statistical method for handwritten chinese character recognition, in: Proceedings of the 2002 International Conference on Machine Learning and Cybernetics, 2002, vol. 1, 2002, pp. 252–255.
3. [Z. Zheng, B. Padmanabhan, Constructing ensembles from data envelopment](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0635) [analysis, INFORMS Journal on Computing 19 (4) (2007) 486–496](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0635).
4. [D. Wolpert, Stacked generalization, Neural Networks 5 (1992) 241–259](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0640).
5. [Y. Huang, C. Suen, A method of combining multiple experts for the](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0645) [recognition of unconstrained handwritten numerals, IEEE Transactions on](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0645) [Pattern Analysis and Machine Intelligence 17 (1) (1995) 90–94](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0645).
6. [M.M. Gaber, A. Zaslavsky, S. Krishnaswamy, Mining data streams: a review,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0650) [SIGMOD Record 34 (2) (2005) 18–26](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0650).
7. [A. Patcha, J.-M. Park, An overview of anomaly detection techniques: existing](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0655) [solutions and latest technological trends, Computer Network 51 (12) (2007)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0655) [3448–3470](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0655).
8. [M.M. Black, R.J. Hickey, Classiﬁcation of customer call data in the presence of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0660) [concept drift and noise, in: Proceedings of the First International Conference](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0660) [on Computing in an Imperfect World, Soft-Ware 2002,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0660) [Springer-Verlag,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0660) [London, UK, 2002, pp. 74–87](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0660).
9. [H. Wang, W. Fan, P.S. Yu, J. Han, Mining concept-drifting data streams using](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0665) [ensemble classiﬁers, in: Proceedings of the Ninth ACM SIGKDD International](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0665) [Conference on Knowledge Discovery and Data Mining, KDD ’03,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0665) [ACM, New](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0665) [York, NY, USA, 2003, pp. 226–235](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0665).
10. M.M. Gaber, P.S. Yu, Classiﬁcation of changes in evolving data streams using online clustering result deviation, in: Proc. Of International Workshop on Knowledge Discovery in Data Streams, 2006.
11. [M. Markou, S. Singh, Novelty detection: a review – Part 1: Statistical](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0670) [approaches, Signal Process 83 (12) (2003) 2481–2497](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0670).
12. [M. Salganicoff, Density-adaptive learning and forgetting, in: Machine](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0675) [Learning: Proceedings of the Tenth Annual Conference,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0675) [Morgan Kaufmann,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0675) [San Francisco, CA, 1993](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0675).
13. [R. Klinkenberg, T. Joachims, Detecting concept drift with support vector](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0680) [machines, in: Proceedings of the Seventeenth International Conference on](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0680) [Machine Learning, ICML ’00, Morgan Kaufmann Publishers Inc., San Francisco,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0680) [CA, USA, 2000, pp. 487–494](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0680).
14. M. Baena-Garc´ıa, J. del Campo-Ávila, R. Fidalgo, A. Bifet, R. Gavaldá, R. Morales-Bueno, Early drift detection method, in: Fourth International Workshop on Knowledge Discovery from Data Streams, 2006.
15. [I. Zliobaite, Change with delayed labeling: when is it detectable?, in:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0685) [Proceedings of the 2010 IEEE International Conference on Data Mining](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0685) [Workshops, ICD-MW ’10, IEEE Computer Society, Washington, DC, USA, 2010,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0685) [pp 843–850](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0685).
16. [G. Giacinto, F. Roli, L. Bruzzone, Combination of neural and statistical](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0690) [algorithms for supervised classiﬁcation of remote-sensing images, Pattern](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0690) [Recognition Letters 21 (5) (2000) 385–397](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0690).
17. [V. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J. Rigol-Sanchez,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0695) [An assessment of the effectiveness of a random forest classiﬁer for land-cover](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0695) [classiﬁcation, ISPRS Journal of Photogrammetry and Remote Sensing 67 (0)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0695) [(2012) 93–104](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0695).
18. [P. Gislason, J. Benediktsson, J. Sveinsson, Random forests for land cover](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0700) [classiﬁcation, Pattern Recognition Letters 27 (4) (2006) 294–300](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0700).
19. [J.-W. Chan, D. Paelinckx, Evaluation of random forest and Adaboost tree-](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0705) [based ensemble classiﬁcation and spectral band selection for ecotope](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0705) [mapping using airborne hyperspectral imagery, Remote Sensing of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0705) [Environment 112 (6) (2008) 2999–3011](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0705).
20. [J. Peters, N. Verhoest, R. Samson, M. Meirvenne, L. Cockx, B. Baets,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0710) [Uncertainty propagation in vegetation distribution models based on](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0710) [ensemble classiﬁers, Ecological Modelling 220 (6) (2009) 791–804](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0710).
21. [M. Han, X. Zhu, W. Yao, Remote sensing image classiﬁcation based on neural](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0715) [network ensemble algorithm, Neurocomputing 78 (1) (2012) 133–138](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0715).
22. [B. Waske, M. Braun, Classiﬁer ensembles for land cover mapping using](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0720) [multitemporal SAR imagery, ISPRS Journal of Photogrammetry and Remote](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0720) [Sensing 64 (5) (2009) 450–457 (theme Issue: Mapping with SAR: Techniques](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0720) [and Applications)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0720).
23. [B. Waske, S. van der Linden, C. Oldenburg, B. Jakimow, A. Rabe, P. Hostert,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0725) [imageRF – a user-oriented implementation for remote sensing image analysis](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0725) [with random forests, Environmental Modelling & Software 35 (0) (2012)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0725) [192–193](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0725).
24. [U. Maulik, D. Chakraborty, A self-trained ensemble with semisupervised](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0730) [SVM: an application to pixel classiﬁcation of remote sensing imagery, Pattern](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0730) [Recognition 44 (3) (2011) 615–623](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0730).
25. [A. Henriques, A. Doria-Neto, R. Amaral, Classiﬁcation of multispectral images](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0735) [in coral environments using a hybrid of classiﬁer ensembles,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0735) [Neurocomputing 73 (7–9) (2010) 1256–1264](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0735).
26. [Y. Maghsoudi, M. Collins, D. Leckie, Polarimetric classiﬁcation of boreal forest](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0740) [using nonparametric feature selection and multiple classiﬁers, International](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0740) [Journal of Applied Earth Observation and Geoinformation 19 (0) (2012) 139–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0740) [150](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0740).
27. [L. Bruzzone, R. Cossu, G. Vernazza, Combining parametric and non-](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0745) [parametric algorithms for a partially unsupervised classiﬁcation of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0745) [multitemporal remote-sensing images, Information Fusion 3 (4) (2002)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0745) [289–297](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0745).
28. [L. Bruzzone, R. Cossu, G. Vernazza, Detection of land-cover transitions by](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0750) [combining multidate classiﬁers, Pattern Recognition Letters 25 (13) (2004)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0750) [1491–1500](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0750).
29. [P. Du, S. Liu, J. Xia, Y. Zhao, Information fusion techniques for change](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0755) [detection from multi-temporal remote sensing images, Information Fusion](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0755) [14 (1) (2013) 19–27](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0755).
30. [P. Arun-Raj-Kumar, S. Selvakumar, Distributed denial of service attack](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0760) [detection using an ensemble of neural classiﬁer, Computer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0760) [Communications 34 (11) (2011) 1328–1341](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0760).
31. [P. Kumar, S. Selvakumar, Detection of distributed denial of service attacks](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0765) [using an ensemble of adaptive and hybrid neuro-fuzzy systems, Computer](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0765) [Communications (0) (2012)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0765).
32. [A. Shabtai, R. Moskovitch, Y. Elovici, C. Glezer, Detection of malicious code by](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0770) [applying machine learning classiﬁers on static features: a state-of-the-art](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0770) [survey, Information Security Technical Report 14 (1) (2009) 16–29](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0770).
33. [M. Locasto, K. Wang, A. Keromytis, S. Stolfo, Flips: hybrid adaptive intrusion](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0775) [prevention, in: Proceedings of the 8th International Conference on Recent](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0775) [Advances in Intrusion Detection, RAID’05,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0775) [Springer-Verlag, Berlin,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0775) [Heidelberg, 2006, pp. 82–101](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0775).
34. [K. Wang, G. Cretu, S. Stolfo, Anomalous payload-based worm detection and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0780) [signature generation, in: Proceedings of the 8th International Conference on](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0780) [Recent Advances in Intrusion Detection, RAID’05,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0780) [Springer-Verlag, Berlin,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0780) [Heidelberg, 2006, pp. 227–246](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0780).
35. [S. Peddabachigari, A. Abraham, C. Grosan, J. Thomas, Modeling intrusion](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0785) [detection system using hybrid intelligent systems, Journal of Network and](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0785) [Computer Applications 30 (1) (2007) 114–132](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0785).
36. [D.-I. Curiac, C. Volosencu, Ensemble based sensing anomaly detection in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0790) [wireless sensor networks, Expert Systems with Applications 39 (10) (2012)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0790) [9087–9096](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0790).
37. [R.J. Bolton, D.J. Hand, Statistical fraud detection: a review, Statistical Science](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0795) [17 (3) (2002) 235–255](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0795).
38. [F. Louzada, A. Ara, Bagging k-dependence probabilistic networks: an](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0800) [alternative powerful fraud detection tool, Expert Systems with Applications](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0800) [39 (14) (2012) 11583–11592](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0800).
39. [S. Bhattacharyya, S. Jha, K. Tharakunnel, J. Westland, Data mining for credit](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0805) [card fraud: a comparative study, Decision Support Systems 50 (3) (2011)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0805) [602–613](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0805).
40. [L. Yu, W. Yue, S. Wang, K. Lai, Support vector machine based multiagent](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0810) [ensemble learning for credit risk evaluation, Expert Systems with](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0810) [Applications 37 (2) (2010) 1351–1360](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0810).
41. [Y. Kim, S. Sohn, Stock fraud detection using peer group analysis, Expert](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0815) [Systems with Applications 39 (10) (2012) 8986–8992](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0815).
42. [B. Twala, Multiple classiﬁer application to credit risk assessment, Expert](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0820) [Systems with Applications 37 (4) (2010) 3326–3336](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0820).
43. [S. Finlay, Multiple classiﬁer architectures and their application to credit risk](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0825) [assessment, European Journal of Operational Research 210 (2) (2011) 368–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0825) [378](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0825).
44. [G. Wang, J. Ma, A hybrid ensemble approach for enterprise credit risk](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0830) [assessment based on support vector machine, Expert Systems with](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0830) [Applications 39 (5) (2012) 5325–5331](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0830).
45. [M. Kim, D. Kang, Classiﬁers selection in ensembles using genetic algorithms](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0835) [for bankruptcy prediction, Expert Systems with Applications 39 (10) (2012)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0835) [9308–9314](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0835).
46. [P. Ravisankar, V. Ravi, I. Bose, Failure prediction of dotcom companies using](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0840) [neural network–genetic programming hybrids, Information Sciences 180 (8)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0840) [(2010) 1257–1267](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0840).
47. [P. Ravisankar, V. Ravi, G. Rao, I. Bose, Detection of ﬁnancial statement fraud](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0845) [and feature selection using data mining techniques, Decision Support](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0845) [Systems 50 (2) (2011) 491–500](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0845).

*M. Wo´zniak et al. / Information Fusion 16 (2014) 3–17* 17

1. [C. Tsai, Combining cluster analysis with classiﬁer ensembles to predict](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0850) [ﬁnancial distress, Information Fusion (0) (2011)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0850).
2. [Y. Peng, G. Wang, G. Kou, Y. Shi, An empirical study of classiﬁcation algorithm](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0855) [evaluation for ﬁnancial risk prediction, Applied Soft Computing 11 (2) (2011)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0855) [2906–2915](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0855).
3. [V. Ravi, H. Kurniawan, P. Nwee-Kok-Thai, P. Ravi-Kumar, Soft computing](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0860) [system for bank performance prediction, Applied Soft Computing 8 (1) (2008)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0860) [305–315](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0860).
4. [H. Zhao, A. Sinha, W. Ge, Effects of feature construction on classiﬁcation](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0865) [performance: an empirical study in bank failure prediction, Expert Systems](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0865) [with Applications 36 (2, Part 2) (2009) 2633–2644](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0865).
5. [K. Aral, H. Guvenir, I. Sabuncuoglu, A. Akar, A prescription fraud detection](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0870) [model, Computer Methods and Programs in Biomedicine 106 (1) (2012) 37–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0870) [46](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0870).
6. [I. Christou, M. Bakopoulos, T. Dimitriou, E. Amolochitis, S. Tsekeridou, C.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0875) [Dimitriadis, Detecting fraud in online games of chance and lotteries, Expert](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0875) [Systems with Applications 38 (10) (2011) 13158–13169](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0875).
7. [H. Farvaresh, M. Sepehri, A data mining framework for detecting subscription](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0880) [fraud in telecommunication, Engineering Applications of Artiﬁcial](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0880) [Intelligence 24 (1) (2011) 182–194](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0880).
8. [L. Subelj, S. Furlan, M. Bajec, An expert system for detecting automobile](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0885) [insurance fraud using social network analysis, Expert Systems with](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0885) [Applications 38 (1) (2011) 1039–1052](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0885).
9. [A.X. Garg, N.K.J. Adhikari, H. McDonald, M.P. Rosas-Arellano, P.J. Devereaux, J.](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0890) [Beyene, J. Sam, R.B. Haynes, Effects of computerized clinical decision support](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0890) [systems on practitioner performance and patient outcomes: a systematic](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0890) [review, Journal of the American Medical Association 293 (10) (2005) 1223–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0890) [1238](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0890).
10. [J. Eom, S. Kim, B. Zhang, AptaCDSS-E: a classiﬁer ensemble-based clinical](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0895) [decision support system for cardiovascular disease level prediction, Expert](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0895) [Systems with Applications 34 (4) (2008) 2465–2479](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0895).
11. [R. Das, I. Turkoglu, A. Sengur, Effective diagnosis of heart disease through](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0900) [neural networks ensembles, Expert Systems with Applications 36 (4) (2009)](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0900) [7675–7680](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0900).
12. [R. Das, I. Turkoglu, A. Sengur, Diagnosis of valvular heart disease through](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0905) [neural networks ensembles, Computer Methods and Programs in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0905) [Biomedicine 93 (2) (2009) 185–191](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0905).
13. [W. Baxt, Improving the accuracy of an artiﬁcial neural network using](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0910) [multiple differently trained networks, Neural Computation 4 (5) (1992) 772–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0910) [780](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0910).
14. [X. Zhang, J. Mesirov, D. Waltz, Hybrid system for protein secondary structure](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0915) [prediction, Journal of Molecular Biology 225 (4) (1992) 1049–1063](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0915).
15. [L. Nanni, Ensemble of classiﬁers for protein fold recognition,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0920) [Neurocomputing 69 (7) (2006) 850–853](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0920).
16. [T. Yang, V. Kecman, L. Cao, C. Zhang, J.Z. Huang, Margin-based ensemble](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0925) [classiﬁer for protein fold recognition, Expert Systems with Applications 38](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0925) [(10) (2011) 12348–12355](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0925).
17. [A. Savio, M. Garcia-Sebastian, D. Chyzyk, C. Hernandez, M. Graña, A. Sistiaga,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0930)

[A.L. de Munain, J. Villanua, Neurocognitive disorder detection based on](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0930) [feature vectors extracted from VBM analysis of structural MRI, Computers in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0930) [Biology and Medicine 41 (8) (2011) 600–610](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0930).

1. [D. Chyzhyk, M. Graña, A. Savio, J. Maiora, Hybrid dendritic computing with](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0935) [kernel-LICA applied to alzheimer’s disease detection in MRI, Neurocomputing](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0935) [75 (1) (2012) 72–77](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0935).
2. [L. Kuncheva, J. Rodriguez, Classiﬁer ensembles for fMRI data analysis: an](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0940) [experiment, Magnetic Resonance Imaging 28 (4) (2010) 583–593](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0940).
3. [C. Plumpton, L. Kuncheva, N. Oosterhof, S. Johnston, Naive random subspace](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0945) [ensemble with linear classiﬁers for real-time classiﬁcation of fMRI data,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0945) [Pattern Recognition 45 (6) (2012) 2101–2108](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0945).
4. [C. Cabral, M. Silveira, P. Figueiredo, Decoding visual brain states from fMRI](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0950) [using an ensemble of classiﬁers, Pattern Recognition 45 (6) (2012) 2064–](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0950) [2074](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0950).
5. [G. Adomavicius, R. Sankaranarayanan, S. Sen, A. Tuzhilin, Incorporating](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0955) [contextual information in recommender systems using a multidimensional](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0955) [approach, ACM Transactions Information Systems 23 (1) (2005) 103–145](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0955).
6. [J. Konstan, J. Riedl, How online merchants predict your preferences and prod](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0960) [you to purchase, IEEE Spectrum 49 (10) (2012) 48–56](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0960).
7. [R. Burke, Hybrid recommender systems: survey and experiments, User](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0965) [Modeling and User-Adapted Interaction 12 (4) (2002) 331–370](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0965).
8. [M. Balabanovic´, Y. Shoham, Fab: content-based, collaborative](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0970) [recommendation, Communications of the ACM 40 (3) (1997) 66–72](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0970).
9. [M.J. Pazzani, A framework for collaborative, content-based and demographic](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0975) [ﬁltering, Artiﬁcial Intelligence Review 13 (5–6) (1999) 393–408](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0975).
10. [M. Jahrer, A. Töscher, R. Legenstein, Combining predictions for accurate](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0980) [recommender systems, in: Proceedings of the 16th ACM SIGKDD](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0980) [International Conference on Knowledge Discovery and Data Mining, KDD](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0980)

[’10,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0980) [ACM, New York, NY, USA, 2010, pp. 693–702](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0980).

1. [C. Porcel, A. Tejeda-Lorente, M. Mart](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0985)´ı[nez, E. Herrera-Viedma, A hybrid](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0985) [recommender system for the selective dissemination of research resources in](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0985) [a technology transfer ofﬁce, Information Sciences 184 (1) (2012) 1–19](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0985).
2. [M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, M. Sartin,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0990) [Combining content-based and collaborative ﬁlters in an online newspaper,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0990) [in: Proceedings of the ACM SIGIR ’99 Workshop on Recommender Systems:](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0990) [Algorithms and Evaluation,](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0990) [ACM, 1999](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0990).
3. [D. Billsus, M. Pazzani, User modeling for adaptive news access, User Modeling](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0995) [and User-Adapted Interaction 10 (2–3) (2000) 147–180](http://refhub.elsevier.com/S1566-2535(13)00047-X/h0995).
4. T. Tran, R. Cohen, Hybrid recommender systems for electronic commerce, in: Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, AAAI Technical Report WS-00-04, AAAI Press, Menlo Park, CA, 2000, pp. 78–83.
5. [M. Kunaver, T. Pozrl, M. Pogacnik, J. Tasic, Optimisation of combined](http://refhub.elsevier.com/S1566-2535(13)00047-X/h1000) [collaborative recommender systems, AEU – International Journal of](http://refhub.elsevier.com/S1566-2535(13)00047-X/h1000) [Electronics and Communications 61 (7) (2007) 433–443](http://refhub.elsevier.com/S1566-2535(13)00047-X/h1000).
6. [M. Goksedef, S. Gundoz-Oguducu, Combination of web page recommender](http://refhub.elsevier.com/S1566-2535(13)00047-X/h1005) [systems, Expert Systems with Applications 37 (4) (2010) 2911–2922](http://refhub.elsevier.com/S1566-2535(13)00047-X/h1005).