Algorithms **2014**, *7*, 405-417; doi:10.3390/a7030405

OPEN ACCESS

algorithms

ISSN 1999-4893

www.mdpi.com/journal/algorithms

Article

Seminal Quality Prediction Using Clustering-Based Decision Forests

Hong Wang, Qingsong Xu and Lifeng Zhou *

School of Mathematics & Statistics, Central South University, Changsha, Hunan, 410075, China;
E-Mails: wh@csu.edu.cn (H.W.); qxsu@csu.edu.cn (Q.X.)

* Author to whom correspondence should be addressed; E-Mail: lfzhou@csu.edu.cn;
Tel.: +86-731-88660140.

Received: 14 May 2014; in revised form: 23 June 2014 / Accepted: 29 July 2014 / Published: 11 August 2014

Abstract: Prediction of seminal quality with statistical learning tools is an emerging methodology in decision support systems in biomedical engineering and is very useful in early diagnosis of seminal patients and selection of semen donors candidates. However, as is common in medical diagnosis, seminal quality prediction faces the class imbalance problem. In this paper, we propose a novel supervised ensemble learning approach, namely Clustering-Based Decision Forests, to tackle unbalanced class learning problem in seminal quality prediction. Experiment results on real fertility diagnosis dataset have shown that Clustering-Based Decision Forests outperforms decision tree, Support Vector Machines, random forests, multilayer perceptron neural networks and logistic regression by a noticeable margin. Clustering-Based Decision Forests can also be used to evaluate variables' importance and the top five important factors that may affect semen concentration obtained in this study are age, serious trauma, sitting time, the season when the semen sample is produced, and high fevers in the last year. The findings could be helpful in explaining seminal concentration problems in infertile males or pre-screening semen donor candidates.

Keywords: seminal prediction; imbalanced learning; variable importance

1. Introduction

Seminal quality is a good predictor of the male potential fertility and causes 25% or more of the infertility problem [1]. Predicting the results of the semen analysis based on the data such as environmental factors or life habits is very useful in early diagnosis of seminal disordered patients, selection of semen donors candidates and the prioritization of further infertility treatment [2–5].

To get a quick estimate of the seminal profile of the patients based on data of life style or personal habits, machine learning and data mining tools have been applied in decision support systems in biomedical engineering [4,5]. In the pioneering work by David Gil *et al.* [4], three kinds of machine learning tools including decision trees (DT), multilayer perceptron (MLP) artificial neural network and support vector machines (SVM) were compared. In a most recent work [5], a neural network approach based on MLP was further proposed to estimate semen parameters.

As is common in bioinformatics [6] and medical diagnosis [7], most available seminal quality prediction data also have the class imbalance problem, *i.e.*, the number of negative (majority) observations far outnumbers the number of positive (minority) observations. Because of this skewed distribution in data, many standard learning algorithms often perform poorly on the minority class [8]. Due to complexity of the problem itself, the data imbalance and its impact on prediction accuracy in seminal quality need further study.

In this paper, we propose a novel supervised ensemble learning approach, namely Clustering-Based Decision Forests (CBDF), to tackle unbalanced class learning problem in seminal quality prediction. In the proposed algorithm, to alleviate the class imbalance problem, the borderline majority data called Tomek Links [9] observations are first removed from the training data. Then, the remaining majority training data will be further clustered into a number of subsets of majority samples. We use each majority subset in turn and a bagging [10] version of the minority data to train each base classifier within the ensemble. To further achieve ensemble diversity, random subspace [11] technique is then applied to the new training data to obtain diverse training sets for the base learners in the ensemble. Different from random forests [12], base learners (decision trees) within CBDF are aggregated using a weighted majority voting scheme. All these together improve the prediction accuracy in the final ensemble. This is the major contribution of this paper.

Experiment results on the fertility diagnosis dataset [4] have shown that in terms of AUC (Area Under the ROC curve), the proposed algorithm outperforms DT, SVM, RF, MLP and LR by a noticeable margin.

As a second contribution, we provide a measure to identify variable importance based on mean Gini decrease in building decision trees, which may help the interpretation of the model results.

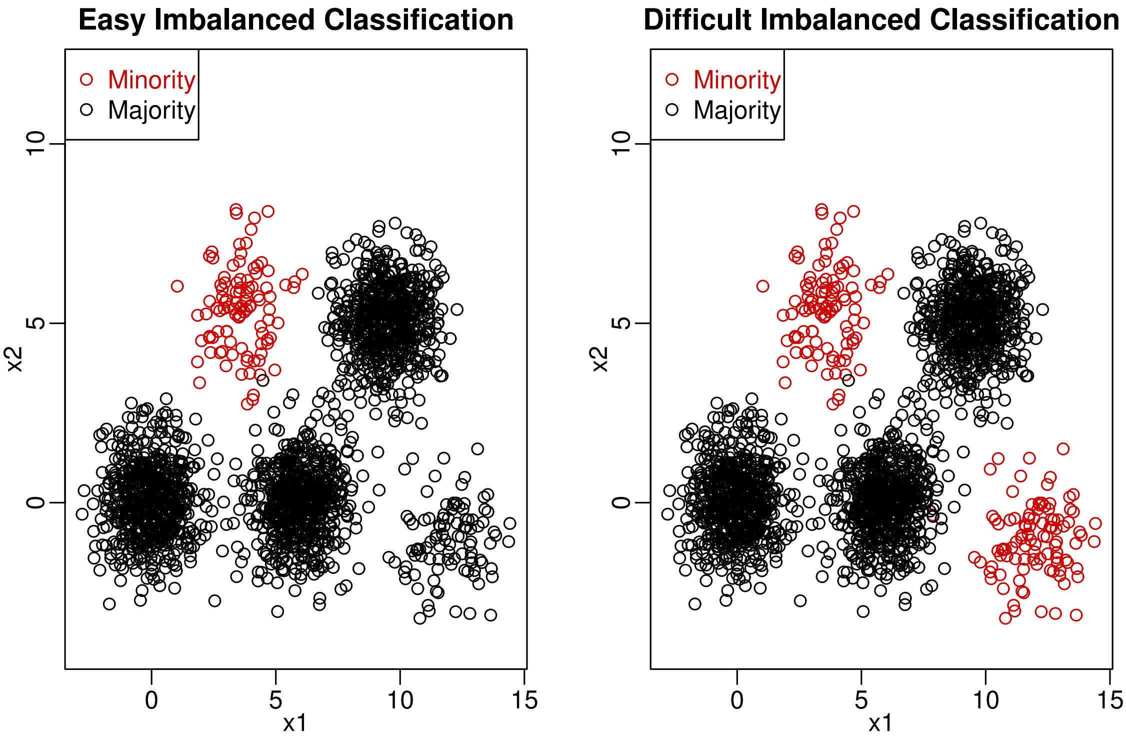
2. Methods

Predicting parameters of seminal quality such as sperm concentration can be viewed as a binary classification problem. Given the training dataset of n labeled observations $D = \{(\mathbf{x}_i, y_i)|i = 1, 2, \cdots, n\}$, where $\mathbf{x} = (x_1, x_2, \cdots, x_p)^T \in X$ is a vector of variables (life habits, environmental factors, *etc.*), $y \in Y = \{0, 1\}$ is the class label (0 for "Normal" sperm concentrations and 1 for "Altered" ones in our case). Our goal is to predict unknown sperm concentration result y on new test variables $\mathbf{x}'$ with the learning model $Y = f(X)$ built upon training data D.

As we point out above, most standard learning algorithms do not perform well on imbalanced datasets, which is frequently met in medical decision making practices. Some classification models might work well in the case of easier imbalance classification problem, as shown in the left panel of

Figure 1. However, when majority class samples overlap with the minority class data due to ambiguous boundary or noisy samples (shown in the right panel of Figure 1), it is rather difficult to find a satisfactory classification model for such data.

Figure 1. Why the majority class samples must be clustered.



Here, we use a three-stage ensemble learning framework called Clustering-Based Decision Forests (CBDF) to tackle such problems. In the first stage, to deal with the class imbalance problem, the majority class data are cleaned using the Tomek Links approach. In the second stage, a majority subset and a bagging minority data are combined together to form a balanced dataset to train each base learner. In the third stage, random subspace method is applied to generate a diverse forest of decision trees. We will elaborate the proposed learning algorithm in the following.

Stage One: Removal of majority Tomek Links. Suppose $\mathbf{x}_i, \mathbf{x}_j$ belong to different classes (*i.e.*, $y_i \neq y_j$), and $d(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between them. A $(\mathbf{x}_i, \mathbf{x}_j)$ pair is called a *Tomek Link* [9] if there is no example $\mathbf{x}_l$, such that $d(\mathbf{x}_i, \mathbf{x}_l) < d(\mathbf{x}_i, \mathbf{x}_j)$ or $d(\mathbf{x}_j, \mathbf{x}_l) < d(\mathbf{x}_i, \mathbf{x}_j)$. In other words, if the nearest neighbors of $\mathbf{x}_i$ and $\mathbf{x}_j$ have different class labels, then these two observations are called a pair of Tomek Links. Tomek Link observations are usually noisy and borderline samples and can be removed to increase prediction accuracy. In unbalanced learning settings, due to scarcity of minority data, for pairs of Tomek Links, we only remove the majority observations and reserve the minority ones.

Stage Two: Balancing the training data. In medical decision making, only a small portion of the samples are minority ones and in the seminal prediction case the altered ones are rare and of interest. Thus, the normal samples are collected to complement the altered ones during the classification process. The chosen semen samples are normal for various reasons and might be further divided into several subsets based on some kind of similarity measure. If a training set is composed of one subset of majority samples and all the minority samples, it would be easy to train a better base classifier using such more balanced training samples.

Based on this consideration, in the proposed algorithm, the majority samples are clustered into k majority subsets. The parameter k is determined by the class imbalance ratio (number of majority samples divided by number of minority examples). One may notice that common clustering algorithms do not necessarily produce k majority subsets of equal size. Thus, to prevent prediction accuracy problems caused by very small sized samples, majority subsets whose sizes are less than a threshold are discarded.

In practice, the number of base classifiers in the ensemble l is usually much larger than k to ensure prediction accuracy. It is not possible to allocate distinct majority sample subsets for all base classifiers. Hence, in building the CBDF ensemble, the majority subset in the current iteration is chosen by cycling through all available majority subsets.

Stage Three: Diversifying the data. To increase data diversity, the minority samples can be generated by a uniformly sampling with replacement method such as bootstrap aggregating (bagging) [10] methods. Then, a random subspace method [11] is applied to all the training data to create further variable diversity in samples.

By excluding noisy majority samples, subsetting majority data and bagging minority ones, and by random subsetting the variable set, the training data for each base classifier become somewhat clean, diverse and more balanced.

In aggregating all base classifiers, different from simple majority voting method in random forests [12], we use the current iteration training data to examine the performance of the base classifier in question. The resulting prediction accuracies will be used to calculate base classifiers' weights in the final learning ensemble. In CBDF, $w[i]$ is computed according to the following function:

$$w[i] = \begin{cases} 0, & auc_i \leq 0.5 \\ (auc_{max} - auc_i)/(auc_{max} - auc_{min}), & auc_{min} \leq auc_i \leq auc_{max} \end{cases} \tag{1}$$

where auc_{max}, auc_{min} are the maximum and minimum values among all AUCs that are greater than 0.5 respectively.

When the goal of medical research is interpretation, calculating variable importance score becomes a key concern as the scores can be used to identify relevant variables and get a deeper understanding of the data. In tree ensembles, decrease of node impurity in constructing the decision tree base classifiers can be used to measure variable importance. As CBDF is an ensemble of decision trees, calculating variable importance is natural byproduct and the importance of variables can be quantified by tree node split criterion (node impurity measures).

In CBDF, the variable importance for variable j in terms of mean Gini decrease is defined by:

$$IMP_j^{Gini} = \frac{1}{l}\sum_{i=1}^{l} IMP_j^{Gini}(t_i) \tag{2}$$

where $IMP_j^{Gini}(t_i)$ denotes the total reduction in impurity of variable j when splitting the nodes for building the i-th decision tree t_i [13].

The pseudo-code of the proposed CBDF algorithm is presented in Algorithm 1.

Algorithm 1 The CBDF Algorithm

```
Input:  $D$, Training Dataset
        $X$, Feature space $x_1, \cdots, x_p$ of samples in $D$
        $F_i$, Feature subset of $X$ used in the $i$-th iteration
        $BaseTree$, Decision tree learning algorithm
        $k$, Number of majority subsets
        $l$, The ensemble size
Procedure:
  Remove majority Tomek Links data in $D$ to form $D'$
  Clustering the majority samples in $D'$ into $k$ majority subsets $Maj_i (i \in 1, \cdots, k)$ and discarding those subsets whose size are less than a threshold.
  Initialize the ensemble $C = \emptyset$
  FOR $(i = 1; i \leq l; i++)$ {
      Generate bootstrap samples $Min_i$ from minority samples in $D'$.
      Select and combine majority subset $Maj_i$ with $Min_i$ to form training set $L_i$
      Generate $L'_i$ by randomly drawing feature subset $F_i$ of $F$ from $L_i$
      Train the base learner on $L'_i$, let $C_i = BaseTree(L'_i)$
      Use $C_i$ to predict the data not used in training $(D' - L_i)$ and obtain the corresponding AUC value $auc_i$
      $C = C \bigcup C_i$ }
  Calculate weights $w[i]$s for all classifiers based on their $auc_i$s according to Equation (1)
```

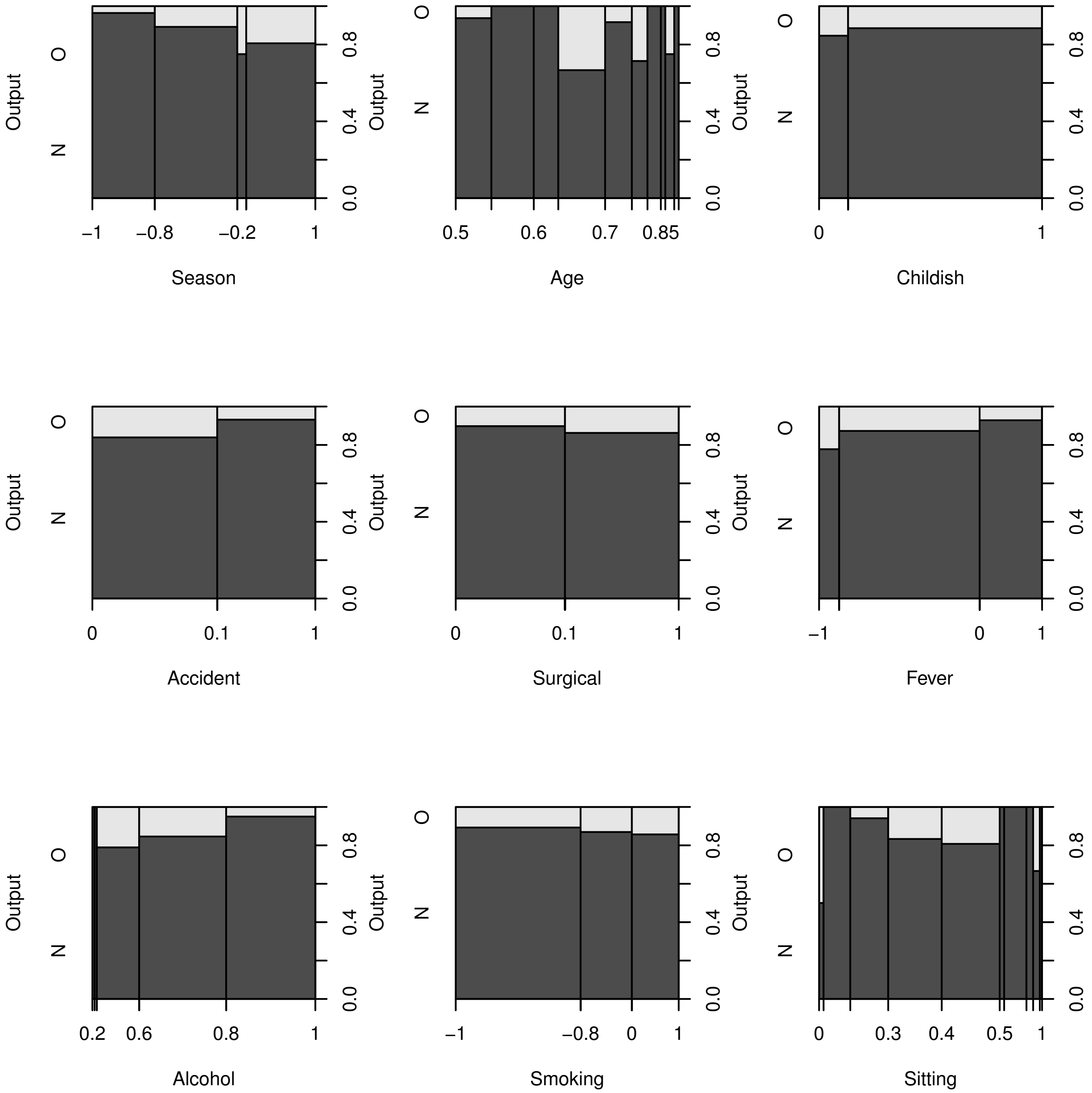
Output: The learning ensemble CBDF C. In prediction, a sample $(\mathbf{x}, y)$ is assigned with class label y^* as the one receiving the weighted majority of the votes:

$$y^* = \arg\max_{y} \sum_{C_i \in C} w[i] \times C_i(\mathbf{x}, y)$$

3. Results

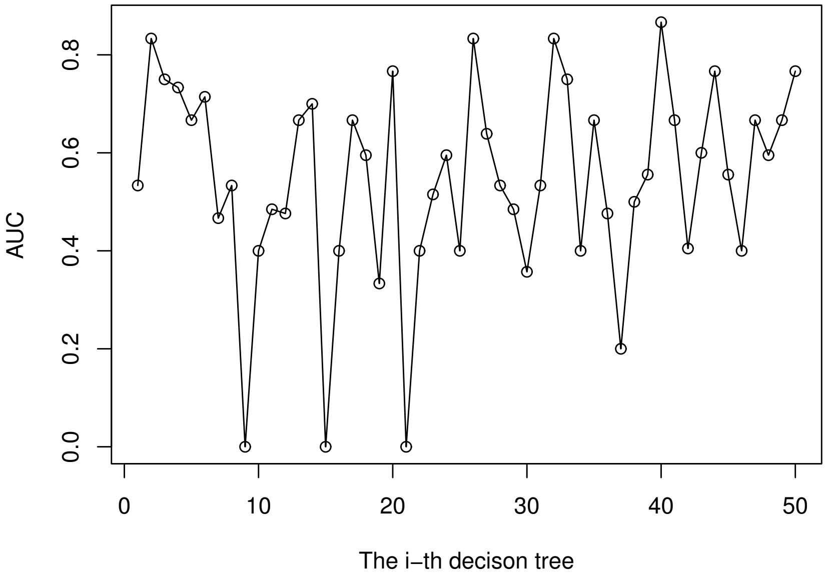
We evaluate the performance of our algorithm on real-world semen quality dataset, *i.e.*, the Fertility Dataset based on [5], which is publicly available on UCI data repository [14]. The Fertility dataset includes a labeled dataset of 100 anonymous young healthy university students between 18 and 36 years old. Each instance in the dataset contains 9 normalized variables about life habit, health status and a semen quality (concentration) result labeled as "normal" and "altered". Fertility dataset is highly unbalanced with a class ratio of almost 7:1 and there are 88 normal and 12 altered instances in the whole dataset. Histograms of all 9 variables regarding semen concentration results (Output) are given in Figure 2. For standard classification problems, the common evaluation criterion is prediction accuracy. However, in unbalanced learning, area under the receiver operating characteristic (ROC) curve, or simply AUC is preferred as it avoids subjectivity in the cutoff selection process [15–17].

Figure 2. Histograms of all variables.



We have obtained the AUC metric estimates via a five-fold cross-validation. In a five-fold cross-validation, the dataset is randomly divided into five equal disjoint subsets, each subset containing 1/5 data of the original dataset. In each fold, the current fold (testing set) is used for validation of the learning model trained on the other folds (training set). Results from all five folds are then averaged to produce a mean AUC estimation. The proposed CBDF is implemented in R [18] with CART (Classification and Regression Tree) [13] as the base learner. Performance of each base learner (CART decision tree in this study) in the ensemble is usually diverse. To illustrate this point, we randomly choose the experiment data in one of the five folds and the performance of all decision trees in terms of AUC is presented in Figure 3.

Figure 3. Performance of base learners in terms of AUC.

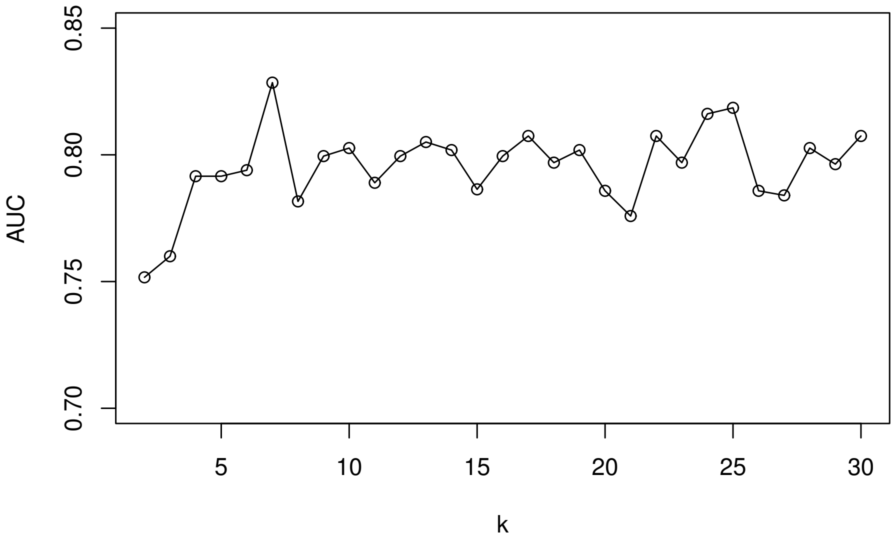


3.1. Parameter Tuning

There are several parameters to tune in the proposed CBDF algorithm, for example, the ensemble size l, the size of feature subset F_i and the number of majority subsets k. As is common for all ensemble algorithms, a larger ensemble size means better prediction accuracy and also a longer time in training and testing. In CBDF, the default ensemble size l is set to 50 for a trade-off between performance and efficiency. Similar to parameter m_{try} in random forests, the size of feature subset F_i is also set to $\sqrt{p}$ ($p = 9$ for the Fertility Dataset) for classification problems.

The number of majority subsets k also plays an important role in the proposed algorithm and we will show how the performance of CBDF will change with different values of k. A five-fold test is run for CBDF with each k in $[2, 30]$ and the mean AUC values are plotted against each k in the following Figure 4.

Figure 4. Sensitivity of parameter k with performance.



As shown in Figure 4, a sharp increase in CBDF's performance is first observed when $k < 4$. However, when $k \geq 4$, CBDF is not very sensitive to the parameter k. The mean AUC of CBDF reaches a maximum when k = 7, which is very close to the class imbalance ratio 7.3. Thus, in the following experiments, the parameter k is set to the class imbalance ratio of each fold adaptively.

3.2. Comparison Study

First, we compare the proposed CBDF algorithm with popular supervised classification algorithms including CART, Random Forest (RF), Support Vector Machines (SVM), Multilayer Perceptron (MLP) neural networks and Logistic Regression (LR). Comparison with these models are conducted with the corresponding "rpart", "randomForest", "e1071", "RSNNS" and "stats" packages in R.

In our experiments, we want all methods to have the same opportunities to achieve their best results, thus the default settings in respective R packages are adopted: for CART, the split criteria is set to Gini; for RF, number of trees $ntree$ = 500 and split variables $mtry$ = 3; for SVM, a radial kernel is applied and γ = 1/9; for MLP, the number of units in the hidden layer $size$ = 5, and maximum of iterations, $maxit$ = 100; for LR, the error distribution and link function $family$ = $gaussian$. The detailed experiment results are given in Table 1.

Table 1. Performance comparison in terms of AUC.

Fold	**1**	**2**	**3**	**4**	**5**
CBDF	0.7895	**0.6842**	**0.9444**	0.6905	**0.9167**
CART	0.7895	**0.6842**	0.5278	0.6190	0.5278
RF	0.4210	0.6316	**0.9444**	0.6786	0.75
SVM	0.4210	0.4210	0.7778	**0.7024**	0.8333
MLP	**1**	0.7368	0.4167	0.2619	0.5833
LR	0.2105	0.3684	0.6667	0.6548	0.8889

In Table 1, the highest AUC scores in each fold are highlighted in bold face. From Table 1, we find that CBDF takes the first place three times in five folds, and it beats other learning algorithms with a mean AUC of 0.8051. The mean AUC scores of RF, SVM, CART, MLP and LR are 0.6851, 0.6311, 0.6296, 0.5997 and 0.5579, respectively.

In the following, we will compare CBDF with other commonly used balancing techniques, namely the representing under-sampling approach of deletion majority Tomek Links (TL), the representing over-sampling approach SMOTE (SM) [17]. As CBDF is an ensemble of CART decision trees, it is appropriate to choose another ensemble of CART decision trees as the baseline learner in the comparisons. Thus we will use the most popular ensemble of CART decision trees, Random Forest (RF), as the baseline learner in the following experiments. We preprocess the training data in each fold by only removing the majority Tomek Links (TL), by only generating minority samples by SMOTE (SM), or both. Then an RF model is learnt with the processed training data. Experiment results in all folds are shown in Table 2.

Table 2. Comparison with other balancing methods in terms of AUC.

Fold	**1**	**2**	**3**	**4**	**5**
CBDF	0.7895	0.6842	**0.9444**	**0.6905**	**0.9167**
TL-RF	0.3158	0.5263	**0.9444**	0.6786	0.8056
SM-RF	0.3158	**0.7895**	0.8333	**0.6905**	0.8889
SM+TL-RF	**0.9474**	0.3158	0.5278	0.6190	0.5278

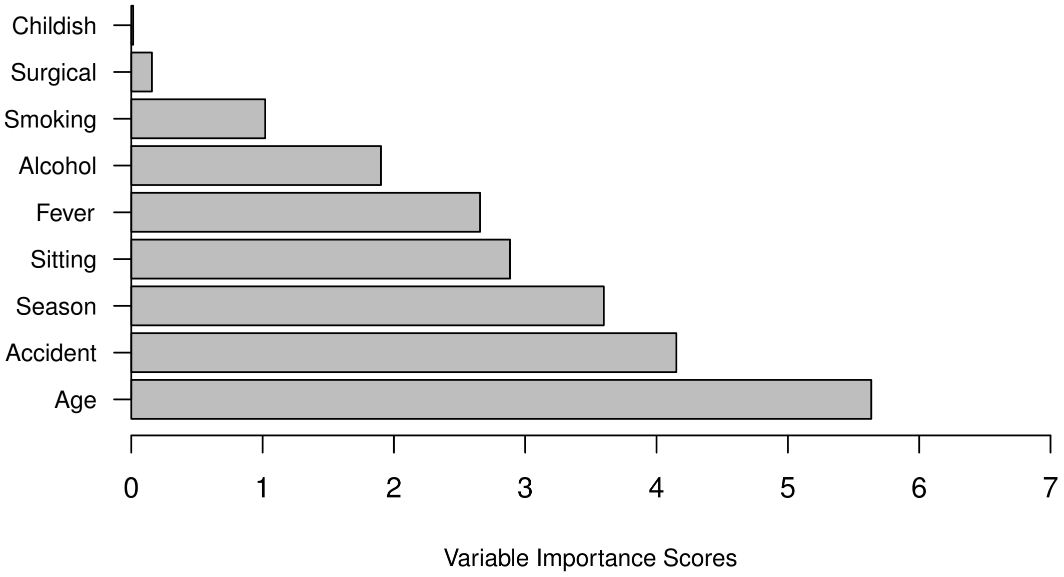
From Table 2, we can find that, CBDF (mean AUC: 0.8051) outperforms TL-RF (mean AUC: 0.6541), SMOTE-RF (mean AUC: 0.7036) and the combination of both (mean AUC: 0.6754) by a large margin. The exclusion of noise Tomek Links, clustering majority samples and bootstrapping minority samples and weighted average of the base classifiers all contributes to CBDF in part.

From Tables 1 and 2, we also notice that RF using SMOTE method also does better than CART, RF, SVM, MLP and LR algorithms. However, it is less stable than CBDF and only gets 0.3158 in Fold 1, which is even worse than a random guess (0.5).

3.3. Variable Importance

As mentioned above, one important feature and advantage of CBDF is that variable importance can be calculated easily. Here, we use CBDF to evaluate all variables and find out which are the most important ones in affecting the seminal quality. We record all the spit impurity information in CART in all five folds and the final mean variable scores are given in Figure 5.

Figure 5. Variable importance scores by CBDF.



From Figure 5, we find that the top five variables affecting semen concentration are Age (age at the time of analysis), Accident (accident or serious trauma), Season (Season in which the analysis was performed), Sitting (Number of hours spent sitting per day), Alcohol (Frequency of alcohol consumption), and Fever (High fevers in the last year). The least important variable found according to Fertility dataset is Childish (Childish diseases).

4. Discussions

In this research, we have compared the performance of CBDF with five popular machine learning algorithms, CART, RF, SVM, MLP and LR, and evaluated their applications in predicting unbalanced semen concentration results. CART, RF, SVM, MLP and LR are highly accurate classifiers and find great applications in various fields but experiment results on Fertility Dataset have shown that they are overtaken by the proposed CBDF in terms of AUC.

Besides CBDF, the performance of CART is the most stable one as AUC scores in all folds is greater than 0.5 but it is much less accurate than CBDF. We notice that MLP, the learning algorithm chosen in [5], gets a perfect classification in fold 1 but its poor performance in other folds demonstrates its instability and makes it an unsatisfactory choice in our research. RF, another ensemble of CART decision trees, is the most similar one to CBDF in performance and could be an alternative when CBDF is not available.

In the experiments, we also justify the contributions of deletion of majority Tomek Links instances, clustering of majority samples and weights placed on base learners. Preprocessing data by only removing majority Tomek Links and/or balancing data by only applying SMOTE all obtain lower AUC scores. These observations suggest that both the noise induced by majority samples and the unbalance due to lack of minority data may degrade the performance of the learning algorithms .

Various studies have suggested that age [19], sexual abstinence [20], cigarette smoking [21], food intake [22] , obesity [23], psychosocial stress [24,25] are all associated with altered semen quality, but conclusions about the extent of the deleterious effects vary widely. In this research, we use CBDF to calculate variables' importance and detect which factors are most relevant to semen concentration.

The most important factor related to semen concentration found in this study is "Age". We find males particularly those older than 24 have a higher risk of semen concentration problem, which agrees with a previous research [19] stating that aging increases risks of diminished semen quality. However, our finding is inconsistent with a previous review suggesting that increased male age is not associated with sperm concentration [25].

From the Fertility Dataset, accident or serious trauma is also highly related to semen quality. This is probably because serious traumas, particularly spinal cord injuries, cause changes to the seminal plasma constituents, bladder management, and the neurogenic impairment to the ejaculatory function [26].

The next three important factors are "Season", "Sitting" and "Fever". These factors are all temperate-related variables and their importance justifies the fact that sperm are heat-sensitive and cannot endure high temperatures. The increase in environment temperature and the influence on testicular temperature may explain for why semen samples obtained in winter are probably better than semen samples collected in other seasons (one-tailed Fisher's exact test p value = 0.0952 in this study). Also, sitting too long may cause an elevation of testicular temperature that may result in poor semen quality, which agrees with [27]. In the same way, having high fevers can adversely affect sperm count, motility and morphology, and this result agrees with [28].

Regarding the contribution of alcohol to semen quality, our research results are consistent with previous studies that moderate alcohol consumption ($\leq$125 g per week in this study) is unlikely to have negative effects on semen quality [21,29]. However, when alcohol consumption grows, it might have detrimental effect on semen quality and this findings is also consistent with [30].

In our study, “Smoking”, “Surgery” (Surgical intervention) and “Childish” (Childish diseases) are the least important factors and sperm concentration of males regarding these factors remain in the normal range, but a clear negative trend is observed in Figure 2. Therefore, men with borderline semen quality who wish to have children should quit smoking, and avoid surgical intervention as much as possible.

5. Conclusions

In this research, a supervised ensemble learning algorithm (*i.e.*, Clustering-Based Decision Forests) is proposed. The proposed algorithm preprocesses the data by removing majority Tomek Links, clustering majority samples into subsets, and bagging the minority samples. Experiment results on real fertility diagnosis dataset have shown that this algorithm outperforms decision trees, support vector machines, random forests, multilayer perceptron neural networks and logistic regression by a noticeable margin.

In this study, we also use Clustering-Based Decision Forests to calculate variables’ importance scores. The resulting top five important factors affecting semen concentration found in this study are age, serious trauma, sitting time, the season when the semen sample is produced, and high fevers in the last year. This may help to explain the seminal concentration problems in infertile males or set conditions to screen semen donor candidates. Case-control studies are needed to further identify the role of these life habits or health factors in male’s semen quality.

Due to lack of data, some possible limitations may exist in this research and the major concern refers to sample size. A larger sample size can increase the confidence in the associations observed. Our future research plans to enlarge the datasets to verify the observed relations.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No.11271374 & No. 61003233). The authors thank all reviewers for their valuable and constructive comments, which greatly improved the quality of this paper.

Author Contributions

The work presented here was carried out in collaboration between all authors. HW,QSX and LFZ defined the research theme; HW and QSX designed the algorithm; LFZ carried out the experiments and analyzed the data; HW and LFZ interpreted the results and wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Templeton, A. Infertility-epidemiology, aetiology and effective management. *Health Bull.* **1995**, *53*, 294–298.
2. Osser, S.; Beckman-Ramirez, A.; Liedholm, P. Semen quality of smoking and non-smoking men in infertile couples in a Swedish population. *Acta Obstet. Gynecol. Scand.* **1992**, *71*, 215–218.
3. Petrelli, G.; Mantovani, A. Environmental risk factors and male fertility and reproduction. *Contraception* **2002**, *65*, 297–300.
4. Gil, D.; Girela, J.L.; de Juan, J.; Gomez-Torres, M.J.; Johnsson, M. Predicting seminal quality with artificial intelligence methods. *Expert Syst. Appl.* **2012**, *39*, 12564–12573.
5. Girela, J.L.; Gil, D.; Johnsson, M.; Gomez-Torres, M.J.; de Juan, J. Semen parameters can be predicted from environmental factors and lifestyle using artificial intelligence methods 1. *Biol. Reprod.* **2013**, *88*, doi:10.1095/?biolreprod.112.104653.
6. Radivojac, P.; Chawla, N.V.; Dunker, A.K.; Obradovic, Z. Classification and knowledge discovery in protein databases. *J. Biomed. Inform.* **2004**, *37*, 224–239.
7. Mazurowski, M.A.; Habas, P.A.; Zurada, J.M.; Lo, J.Y.; Baker, J.A.; Tourassi, G.D. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Netw.* **2008**, *21*, 427–436.
8. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
9. Tomek, I. Two modifications of CNN. *IEEE Trans. Syst. Man Cybern.* **1976**, *6*, 769–772.
10. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140.
11. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
12. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
13. Breiman, L. *Classification and Regression Trees*; CRC Press: New York, NY, USA, 1993.
14. Gil, D.; Girela, J.L. UCI Machine Learning Repository: Fertility Data Set. Available online: http://archive.ics.uci.edu/ml/datasets/Fertility (accessed on 8 November 2013).
15. Huang, J.; Ling, C.X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowle. Data Eng.* **2005**, *17*, 299–310.
16. Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874.
17. Galar, M.; Fernández, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* **2012**, *42*, 463–484.
18. R Core Team. *R: A Language and Environment for Statistical Computing*; Technical Report, ISBN 3-900051-07-0; R Foundation for Statistical Computing: Vienna, Austria, 2013. Available online: http://www. R-project. org (accessed on 8 November 2013).
19. Kidd, S.A.; Eskenazi, B.; Wyrobek, A.J. Effects of male age on semen quality and fertility: A review of the literature. *Fertil. Steril.* **2001**, *75*, 237–248.

20. Levitas, E.; Lunenfeld, E.; Weiss, N.; Friger, M.; Har-Vardi, I.; Koifman, A.; Potashnik, G. Relationship between the duration of sexual abstinence and semen quality: Analysis of 9489 semen samples. *Fertil. Steril.* **2005**, *83*, 1680–1686.
21. Künzle, R.; Mueller, M.D.; Hänggi, W.; Birkhäuser, M.H.; Drescher, H.; Bersinger, N.A. Semen quality of male smokers and nonsmokers in infertile couples. *Fertil. Steril.* **2003**, *79*, 287–291.
22. Mendiola, J.; Torres-Cantero, A.M.; Moreno-Grau, J.M.; Ten, J.; Roca, M.; Moreno-Grau, S.; Bernabeu, R. Food intake and its relationship with semen quality: A case-control study. *Fertil. Steril.* **2009**, *91*, 812–818.
23. Hammoud, A.O.; Gibson, M.; Peterson, C.M.; Meikle, A.W.; Carrell, D.T. Impact of male obesity on infertility: A critical review of the current literature. *Fertil. Steril.* **2008**, *90*, 897–904.
24. Gollenberg, A.L.; Liu, F.; Brazil, C.; Drobnis, E.Z.; Guzick, D.; Overstreet, J.W.; Redmon, J.B.; Sparks, A.; Wang, C.; Swan, S.H. Semen quality in fertile men in relation to psychosocial stress. *Fertil. Steril.* **2010**, *93*, 1104–1111.
25. Li, Y.; Lin, H.; Li, Y.; Cao, J. Association between socio-psycho-behavioral factors and male semen quality: Systematic review and meta-analyses. *Fertil. Steril.* **2011**, *95*, 116–123.
26. Patki, P.; Woodhouse, J.; Hamid, R.; Craggs, M.; Shah, J. Effects of spinal cord injury on semen parameters. *J. Spinal Cord Med.* **2008**, *31*, 27–32.
27. Jung, A.; Schuppe, H.C. Influence of genital heat stress on semen quality in humans. *Andrologia* **2007**, *39*, 203–215.
28. Carlsen, E.; Andersson, A.M.; Petersen, J.H.; Skakkebæk, N.E. History of febrile illness and variation in semen quality. *Hum. Reprod.* **2003**, *18*, 2089–2092.
29. Chia, S.E.; Lim, S.T.A.; Tay, S.K.; Lim, S.T. Factors associated with male infertility: A case-control study of 218 infertile and 240 fertile men. *BJOG* **2000**, *107*, 55–61.
30. Muthusami, K.; Chinnaswamy, P. Effect of chronic alcoholism on male fertility hormones and semen quality. *Fertil. Steril.* **2005**, *84*, 919–924.