

# Feature Selection in Under bagging Based Kernel ELM for Handling Class Imbalance Dataset Classification

Roshani Choudhary<sup>a</sup>, Jayesh Kumar Kashyap<sup>a</sup>, Sanyam Shukla<sup>a</sup> and Manasi Gyanchandani<sup>a</sup>

<sup>a</sup> Maulana Azad National Institute of Technology, Bhopal, India

## Abstract

To learn from class imbalance dataset is a big hurdle in classification. As traditional classifiers are unable to handle class imbalance classification effectively. There are various algorithms developed which can handle the problem of class imbalance. To handle a class imbalance problem different approaches are used like sampling, making modifications at the algorithm level, use of ensembles, and evolutionary techniques. Under Bagging based Kernel ELM i.e. UBKELM is a developed variant of the Extreme Learning Machine (ELM) developed to solve the problem of class imbalance. The UBKELM is an ensemble technique that uses under-sampling to handle the imbalance ratio in the component classifiers. Feature selection in component classifiers when creating the ensemble model is a technique that needs more research. This work proposes a Feature Selection in Under bagging Based Kernel ELM (FS-UBKELM) for handling class imbalance dataset classification. In FS-UBKELM some of the features are removed from every component classifier in term to enhance the performance of classification. For the selection of features, we have used the data complexity method. In term to identify the advancement in performance of the developed method we have compared the outcomes with the other state of developed methods of imbalance classification. The results show the performance improvement in the proposed method significantly on class imbalance classification.

## Keywords

UBKELM, FS-UBKELM, Feature Selection, Data Imbalance, CIL(Class imbalance learning)

## 1. Introduction

Various algorithms are designed for classification like SVM, Neural Network, Decision Tree, Naïve Bayes, etc. There are some factors which effects the performance of the classification problem. The problems which arise in classification are class imbalance, class overlapping, small disjuncts, etc. In real world datasets for classification like dataset for cancer detection, dataset for intrusion detection, fraud detection dataset etc. We can frequently observe the class imbalance problems in these datasets. The class imbalance is a situation when the distribution of the classes is different in a given dataset i.e., the instances in the datasets are present in different proportion. If in a class the no of instance are more than the average number of instances then they are called the majority class and if in a class the no. of instances are less than the average number of instances are classed the minority class. In a class imbalance dataset, the traditional classifiers are unable to perform well for classification, because the results are skewed in favor of the majority class.


Different methods are developed to solve the problem of class imbalance classification. These techniques can be classified as data-level, algorithm-level, ensemble, and evolutionary techniques. To deal with the effect of class imbalance on the classifier, data level methods attempt to balance the class distribution.

---


WCES-2022: Workshop on Control and Embedded Systems, April 22 – 24, 2022, Chennai, India.

EMAIL: [jysksyp@gmail.com](mailto:jysksyp@gmail.com) (Jayesh Kumar Kashyap)

ORCID: 0000-0001-5665-1097 (Jayesh Kumar Kashyap)

 © 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Oversampling and under-sampling are two examples of data level method. The algorithm level methods attempt to change the classifier so that it may be used to classify class imbalances. e.g., cost-sensitive methods. The ensemble methods try to improve the classification performance by combining the results after making multiple classifiers. The problem of class imbalance is also addressed using evolutionary approaches like one-class classification, noise reduction, Universum learning, feature selection, etc. Feature selection is a very popular method for performance enhancement of classification problems, but it is very rarely used for imbalance learning.

Under-Bagging based Kernel ELM (UBKELM) [4], It is a generalized single hidden layer feed-forward neural network (SLFN) built to tackle class imbalance classification and is a variation of Extreme Learning Machine (ELM) [1]. We use a feature selection strategy in the suggested study i.e. in the UBKELM to improve its performance on class imbalance problems.

## **2. Literature Review**

### **2.1. Methods to handle Class Imbalance Learning (CIL)**

There exist various methods to handle the class imbalance classification model. Some of these methods are discussed below:

#### **2.1.1. Data Level Methods**

The data level methods try to balance the proportion of classes in the dataset to reduce the biases of the classifier towards the majority class. These methods include different techniques like over sampling techniques, the under-sampling techniques, and the hybrid sampling techniques. Some popular data sampling methods include the synthetic minority oversampling method (SMOTE) [9], random under-sampling (RUS) [5].

#### **2.1.2. Algorithm Level Modifications**

The algorithm level methods perform the modification in the algorithm itself to make them suitable for class imbalance classification. Some popular algorithmic methods which are used to handle class imbalance learning are weighted ELM (WELM) [3], class specifier ELM (CSELM) [10].

#### **2.1.3. Ensemble Techniques**

Ensemble techniques make multiple classifiers and then combine the result of multiple classifiers to make the final decision. It is taught that a decision made by numerous classifiers is superior to a judgement made by a single classifier. Some of the ensemble techniques to handle class imbalance are Easy-Ensemble and Balance Cascade [5], BWELM [7].

### **2.2. Evolutionary Methods**

There exist some evolutionary techniques which are being used for class imbalance learning like one class classification [12], Universum learning [11], feature selection [6], noise filtering [13] etc.

### **2.3. Data Complexity Analysis**

It is found in the literature [14] that there are various data complexities present in a dataset we use for classification. Some of these data complexity measures effect the performance of classification in

class imbalance datasets. This section discusses some of the data complexity measures.

## 2.4. Fisher's discriminant ratio (F1)

The basic variant of fisher's discriminant ratio (F1) use to compute that if we consider any specific feature then how two classes are separated. [14].

## 2.5. The volume of overlap region (F2)

The volume of overlap region computes the length of coincide range i.e., overlap range normalized by total range's length which contain the distributed values in both the classes, then we can obtain the value of volume of coincide region in two classes in term of product of normalized length of all features which are in overlapping ranges. [14].

## 2.6. Feature efficiency (F3)

Feature efficiency is termed as ratio of all the left points which are segregated by particular feature, The greatest feature efficiency (i.e., the largest % of points separable by utilizing that unique feature) is used as an estimate of overlap for a binary class classification issue. [14].

## 2.7. Kernel Extreme Learning Machine (KELM)

The Extreme Learning Machine (ELM) is a single hidden layer feed-forward neural network (SLFN) with excellent generalization and fast learning speed. Extreme Learning Machine can be used for both regression as well as classification also. ELM was originally proposed with two variants sigmoid node-based ELM and Gaussian kernel-based ELM (KELM) [2]. The KELM outperforms the ELM based on sigmoid nodes. The Gaussian kernel function is used by KELM to map the input data to the feature space, as shown below:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right)$$

The kernel matrix of KELM is given as:

$$\Omega_{ELM} = \mathbf{H}\mathbf{H}^T : \Omega_{ELM_{i,j}} = h(x_i) \cdot h(x_j) = K(\mathbf{x}_i, \mathbf{x}_j)$$

The kernel matrix  $\Omega$  is represented as follow for N number of training instances:

$$\Omega = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \dots & K(x_1, x_N) \\ K(x_2, x_1) & K(x_2, x_2) & \dots & K(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_N, x_1) & K(x_N, x_2) & \dots & K(x_N, x_N) \end{bmatrix}_{N \times N}$$

The following equation can be used to calculate the output of KELM [2]:

$$\mathbf{f}(\mathbf{x}) = \text{sign} \left[ \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T \left( \frac{I}{C} + \Omega_{ELM} \right)^{-1} \mathbf{T} \right]$$

## 2.8. Under-Bagging based Kernel ELM (UBKELM)

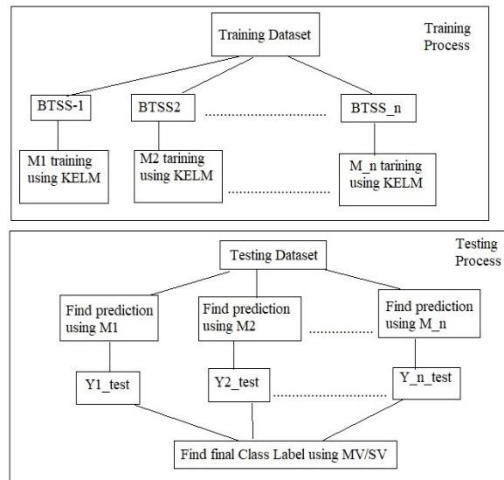
UBKELM [4] proposes an Under-Bagging ensemble which use as the component classifier i.e. use kernelized-ELM. The capabilities of random under-sampling and bagging were inferred using the Under Bagging ensemble.. The training and testing process of UB-KELM is shown in Figure-1.UB-KELM develops several balanced training subsets (BTSS) by randomly under sampling the majority class samples in every training sample. Each BTSS includes all cases of the minority class as well as the equivalent amount of randomly chosen majority class samples. The variety of training subsets i.e., T are dependent on degree of class imbalance which can be obtained by using the following equation:

$$T = \frac{\max(t_k)}{\min(t_k)} \quad t_k \in 1, 2, \dots, m$$

Here,  $t_k$  represent the number of samples belonging to kth class, where m is the number of classes in the dataset.

## 3. Proposed Work

This paper suggests a UBKELM variation to effectively handle a class imbalance classification challenge. The proposed work incorporates the feature selection in the UBKELM algorithm. UBKELM is an ensemble method, the proposed method reduces one feature in every component classification model of UBKELM. For the selection of feature which is to be removed the proposed work used the data complexity analysis. The proposed work used the volume of overlap region(F2) as the measure of complexity for feature reduction in the component classifier. The training and testing process of the developed method is shown in Figure-2.



**Figure 1:** UBKELM training and testing process.

Feature selection UBKELM perform random under-sampling of the majority class samples in each training subset to creates several balanced training subsets (BTSS) by random. Each subset contains all the minority class instances and the same number of randomly selected majority class instances as the minority class. After the creation of balances training subsets, one feature which is having the least volume of overlapping region (F2) value is removed from the corresponding BTSS. The training model is learned using the reduced no. of features for every component model. And number of training subsets i.e., T is identified in the same manner as UBKELM.

## 4. Experimental Setup and Analysis of the Results

In this section, we go over the setup of the tests and the analysis of the results for evaluating the

suggested method's performance.

#### 4.1. Evaluation Matrix

In a class imbalance scenario accuracy cannot be a good measure to check whether the performance is good of a classifier. There are many evaluation matrices which are used for the performance evaluation of classifiers with class imbalance datasets. The Precision, Recall, G-mean, F-measure are some of such measures.

	Predicted	
Actual	<i>True Positive</i>	<i>False Negative</i>
	<i>False Positive</i>	<i>True Negative</i>

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + False\ Negative + True\ Negative}$$

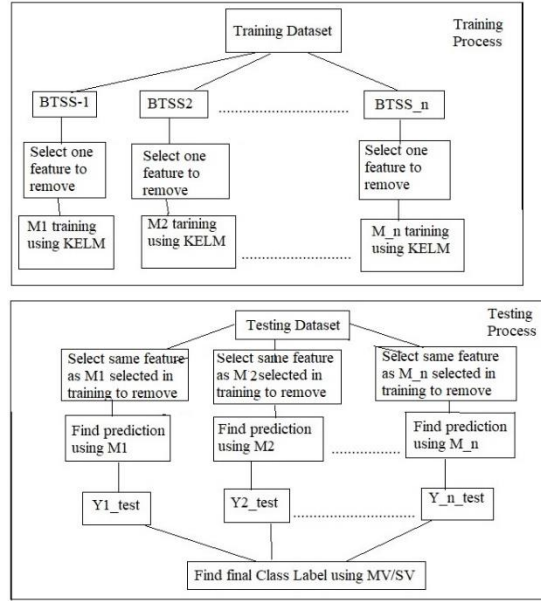
$$G - mean = \sqrt{\frac{True\ Positive}{True\ Positive + False\ Negative} \times \frac{True\ Positive}{True\ Negative + False\ Positive}}$$

#### 4.2. Experimental Setup

We have implemented the code of the proposed FS- UBKELM in Matlab-2017. The proposed method incorporates the comparison with 4 state art methods for class imbalance learning which are WELM, BELM, UBKELM-SV, and UBKELM-MV. The results of these methods are obtained from the work UBKELM.

#### 4.3. Parameter Settings

The proposed method uses the RBF kernel function in which we have to choose the value of the kernel width parameter i.e., represented as  $\sigma$  in the proposed work. The value of  $\sigma$  is tuned using the grid search. The proposed work used the regularized version of ELM i.e., RELM in which the regularization parameter needs to be tuned.



**Figure 2:** FS-UBKELM training and testing process.

In this work the regularization parameter is referred to as  $C$ , the value of  $C$  is tuned using grid search.

The range for tuning  $\sigma$ :

$$2^{-18}, 2^{-17}, \dots, 2^{19}, 2^{20}$$

The range for tuning  $C$ :

$$2^{-18}, 2^{-17}, \dots, 2^{19}, 2^{50}$$

#### 4.4. Result Analysis

For performance evaluation with proposed model the G-mean is used and performance of other methods with proposed method is used for comparison. Keeping other methods in consideration table 1 provide G-mean of proposed methods. In term to show the significant improvement Wilcoxon signed rank test and T-test are carried out and the respective improved results are obtained. While performing T-test, for Null-Hypothesis it returns a test decision that the data into sample space are obtained from normal distribution keeping unknown variance and mean equal to zero., performing the paired sample T-test. If the population distribution doesn't have the value of mean equal to zero then it is named as alternate hypothesis. And if the test fails to reject the null hypothesis at a level of significance of 5%, the result of  $H$  is 1., else the value of  $H$  is 0, the states contain information about the test statistics. In Table-2, we have shown the result of T-test, which confirm that the method which is proposed in this is significantly better compare to all the other methods in consideration. While performing Wilcoxon signed-rank, we get the p-value of paired, two-sided test in term to null hypothesis that the two-population obtained when the median of distribution is zero. To indicate the test decision, we have "H", which return a logical value. If the value of  $H$  is "1", It means that the null hypothesis is reject. Else if the value of  $H$  is "0", It indicate that it is failing to reject the null hypothesis at 5% significance level. The Stats contain information about the test statistic. Table-3 shows the Wilcoxon signed rank test results, which show a considerable performance improvement of the suggested strategy over the other methods under consideration.

**Table 1:** Test G-mean

Datasets	Algorithm					$\sigma$	$C$
	<i>WELM</i>	<i>BELM</i>	<i>UBKELM-SV</i>	<i>UBKELM-MV</i>	<i>FS-UBKELM</i>		
abalone9-18	89.76	90.12	91.07	91.53	92.15	18	36
ecoli-0137vs26	73.65	78.51	81.08	77.74	92.92	8	26

Datasets	Algorithm					$\sigma$	$C$
	WELM	BELM	UBKELM-SV	UBKELM-MV	FS-UBKELM		
ecoli-01_vs_5	88.36	89.36	94.02	93.63	98.62	-6	12
glass016vs2	83.39	84.21	83.39	84.48	87.83	20	42
glass-0123vs456	95.41	94.21	95.45	95.24	96.41	-4	-8
glass2	82.59	85.50	83.26	85.94	87.45	20	44
glass4	91.17	90.34	92.86	92.91	96.69	2	20
haberman	66.26	65.14	66.49	66.70	68.55	10	16
yeast-05679vs4	82.21	80.96	83.45	82.24	83.20	0	-8
yeast-1289vs7	71.41	72.67	74.73	74.28	76.75	10	20
yeast-1458vs7	69.32	69.87	70.15	71.24	71.91	-4	40
yeast-1_vs_7	77.72	77.72	77.90	77.73	80.14	4	8
yeast-2_vs_8	77.89	78.35	81.69	80.39	81.11	10	10
yeast4	84.98	84.74	84.83	85.27	85.92	-2	50

**Table 2:** Statistical T-test

Methods Compared	Statistical Result		
	Stats	P value	H (5%)
WELM vs FS-UBKELM	[-7.495145523745117; -1.864425904826311]	0.003287743	1
BELM vs FS-UBKELM	[-6.239463381338741; -2.037250904375548]	9.38E-04	1
UBKELM_SV vs FS-UBKELM	[-4.582737837292964; -1.026833591278464]	0.004669434	1
UBKELM_MV vs FS-UBKELM	[-5.066495924931664; -0.693075503639768]	0.01378597	1

**Table 3:** Wilcoxon signed rank test

Methods Compared	Statistical Results		
	Signed rank	P value	H (5%)
WELM vs FS-UBKELM	0	1.22E-04	1
BELM Vs FS-UBKELM	0	1.22E-04	1
UBKELM_SV Vs FS-UBKELM	3	6.10E-04	1
UBKELM_MV Vs FS-UBKELM	0	1.22E-04	1

## 5. Conclusion and Future Work

In this paper the proposed method is regarding handling class imbalance learning with the help of combining the feature selection in ensemble method. For the selection of feature, we have used the volume of overlap region (F2). In future we can use other Data complexity measures as well for feature reduction in the training process. Also, the feature selection can be combined with other variant of ELM, which can handle class imbalance like weighted kernel ELM(WKELM), Class-Specific Kernelized ELM(CSKELM), Under bagging Reduced Kernelized Weighted ELM (UBRKWELM).

## 6. References

- [1] Huang, G., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42 , 513-529.
- [2] Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70 , 489-501. *Neural Networks*.
- [3] Zong, W., Huang, G.-B., & Chen, Y. (2013). Weighted extreme learning machine for imbalance learning. *Neurocomputing*, 101 , 229- 242.
- [4] Raghuvanshi, B. S., & Shukla, S. (2019). Class imbalance learning using underbagging based kernelized extreme learning machine. *Neurocomputing*, 329 , 172-187.
- [5] Liu, X., Wu, J., & Zhou, Z. (2009). Exploratory undersampling for class imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39 , 539-550.
- [6] Liu, M., Xu, C., Luo, Y., Xu, C., Wen, Y., & Tao, D. (2018). Cost-sensitive feature selection by optimizing f-measures. *IEEE Transactions on Image Processing*, 27 , 1323-1335.
- [7] Li, K., Kong, X., Lu, Z., Wenyin, L., & Yin, J. (2014). Boosting weighted elm for imbalanced learning. *Neurocomputing*, 128 , 15-21.
- [8] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2012). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42 , 463-484.
- [9] Chawla, N., Bowyer, K., O. Hall, L., & Philip Kegelmeyer, W. (2002). Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res. (JAIR)*, 16 , 321-357.
- [10] Raghuvanshi, B. S., & Shukla, S. (2018a). Class-specific extreme learning machine for handling binary class imbalance problem. *Neural Networks*, 105 , 206-217.
- [11] Zhiqian Qi, Yingjie Tian, Yong Shi, (2014). A nonparallel support vector machine for a classification problem with universum learning. *Journal of Computational and Applied Mathematics*, 263, 288–298.
- [12] Alexandros Iosifidis, Vasileios Mygdalis, Anastasios Tefas, Ioannis Pitas, (2016). One-Class Classification Based on Extreme Learning and Geometric Class Information. *Neural Process Lett.*
- [13] Qi Kang, XiaoShuang Chen, SiSi Li, MengChu Zhou. (2017) A Noise-Filtered Under-Sampling Scheme for Imbalanced Classification. *IEEE TRANSACTIONS ON CYBERNETICS, VOL. 47, NO. 12, 4263-2274*
- [14] J. M. Sotoca, J. S. Sánchez, R. A. Mollineda. (2005) A review of data complexity measures and their applicability to pattern classification problems. *III Taller Nacional de Minería de Datos y Aprendizaje, TAMIDA*, 77-83.
- [15] V.M. Janakiraman, X. Nguyen, D. Assanis, Stochastic gradient based extreme learning machines for stable online learning of advanced combustion engines, *Neurocomputing* 177 (2016) 304–316.
- [16] W. Xiao, J. Zhang, Y. Li, S. Zhang, W. Yang, Class-specific cost regulation extreme learning machine for imbalanced classification, *Neurocomputing* 261(2017) 70–82.
- [17] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches, *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* 42 (4)(2012) 463–484.
- [18] F. Luo, W. Guo, Y. Yu, G. Chen, A multi-label classification algorithm based on kernel extreme learning machine , *Neurocomputing* 260 (2017) 313–320.