



Category: STEM (Science, Technology, Engineering and Mathematics)

ORIGINAL

Data oversampling and feature selection for class imbalanced datasets

Sobremuestreo de datos y selección de características para conjuntos de datos desequilibrados de clase

Krishnakumar V¹  , Sangeetha V²  

¹Department of Computer Science, Periyar University. Salem, India.

²Department of Computer Science, Periyar University Arts and Science College. Pappireddipatti, Dharmapuri, India.

Cite as: Krishnakumar V, Sangeetha V. Data oversampling and feature selection for class imbalanced datasets. Salud, Ciencia y Tecnología - Serie de Conferencias. 2024; 3:935. <https://doi.org/10.56294/sctconf2024935>

Submitted: 10-02-2024

Revised: 01-05-2024

Accepted: 20-06-2024

Published: 21-06-2024

Editor: Dr. William Castillo-González 

ABSTRACT

Introduction: significant advancements and modifications have been implemented in data classification (DC) in the past few decades. Due to their infinite quantity and imbalance, data becomes challenging for classification. The biggest concern in DM (Data Mining) is Class Imbalance (CI). To avoid these issues in recent work proposed map reduce based data parallelization of class imbalanced datasets.

Method: a novel Over Sampling (OS) technique called Minority Oversampling in Kernel Canonical Correlation Adaptive Subspaces (MOKCCAS) has been suggested with the objective to minimize data loss throughout (FSP) Feature Space Projections. This technique takes advantage of the constant Feature Extraction (FE) capability of a version of the ASSOM (Adaptive Subspace Self-Organizing Maps) that is derived from Kernel Canonical Correlation Analysis (KCCA). And in classification, Feature Selection (FS) plays an important role because the acquired dataset might contain large volume of samples, utilizing all features of samples from the dataset for classification will decrease the classifier performance. And then data parallelization will be done by using map reduce framework to solve this computation requirement problem.

Results: then proposes a feature selection model using Mutated whale optimization (MWO) methods and produces features and reduces the time consumption. Finally proposed class balancing model will be tested using uniform distribution based enhanced adaptive neuro fuzzy inference system (UDANFIS). Test outcomes validate the efficiency of the suggested technique by precision, recall, accuracy and Error Rate (ER).

Conclusions: the study subsequently suggests a novel OS approach called MOKCCAS to lessen the loss of data throughout feature space projection.

Keywords: Imbalance Data; Data Mining (DM); Parallelization; Feature Space Projection (FSP); Minority Oversampling (MO); Kernel Canonical Correlation Adaptive Subspaces (KCCAS); Whale Optimization (WO).

RESUMEN

Introducción: se han implementado avances y modificaciones importantes en la clasificación de datos (DC) en las últimas décadas. Debido a su infinita cantidad y desequilibrio, los datos se vuelven difíciles de clasificar. La mayor preocupación en DM (minería de datos) es el desequilibrio de clases (CI). Para evitar estos problemas, en un trabajo reciente se propuso reducir el mapa basado en la paralelización de datos de conjuntos de datos desequilibrados de clases.

Método: se ha sugerido una nueva técnica de sobremuestreo (OS) llamada sobremuestreo minoritario en subespacios adaptativos de correlación canónica del núcleo (MOKCCAS) con el objetivo de minimizar la pérdida de datos en las proyecciones del espacio de características (FSP). Esta técnica aprovecha la capacidad constante de extracción de características (FE) de una versión de ASSOM (mapas autoorganizados adaptativos

del subespacio) que se deriva del análisis de correlación canónica del kernel (KCCA). Y en la clasificación, la selección de características (FS) juega un papel importante porque el conjunto de datos adquirido puede contener un gran volumen de muestras; utilizar todas las características de las muestras del conjunto de datos para la clasificación disminuirá el rendimiento del clasificador. Y luego la paralelización de datos se realizará utilizando el marco de reducción de mapas para resolver este problema de requisitos de cálculo.

Resultados: luego propone un modelo de selección de características utilizando métodos de optimización de ballenas mutadas (MWO), produce características y reduce el consumo de tiempo. Finalmente, el modelo de equilibrio de clases propuesto se probará utilizando un sistema de inferencia neurodifusa adaptativo mejorado basado en distribución uniforme (UDANFIS). Los resultados de la prueba validan la eficiencia de la técnica sugerida mediante precisión, recuperación, exactitud y tasa de error (ER).

Conclusiones: posteriormente, el estudio sugiere un nuevo enfoque de sistema operativo llamado MOKCCAS para reducir la pérdida de datos a lo largo de la proyección del espacio de características.

Palabras clave: Datos de Desequilibrio; Minería de Datos (DM); Paralelización; Proyección del Espacio de Características (FSP); Sobremuestreo Minoritario (MO); Subespacios Adaptativos de Correlación Canónica del Kernel (KCCAS); Optimización de Ballenas (WO).

INTRODUCTION

In several real-time applications, the amount of data continues to expand daily along with different distributions. When there are more specimens in one class in a dataset than in another, it is considered to be extremely divergent. Whereas the minority class has less specimens than the main class, the majority class is employed for recognizing any imbalance in the dataset. Major classes frequently express the specimens in a negative way, while minority classes perform inversely. By class proportions, that may vary from 100 with 1 and 1000 with 1, etc., the number of specimens from the majority class far exceeds specimens from the minority class.^(1,2)

Binary Class (BC) and Multi-Class (MC) datasets both have imbalanced data problems. A BC dataset consists of only two classes, while the MC dataset contains more than two classes. Imbalanced dataset problems occur in many actual domains, like identifying insecure telecom users, learning word pronunciations, identifying oil spills in satellite radar images, (IR) Information Retrieval, Text Classification (TC), filtering tasks, identifying false phone calls, and most crucially, healthcare diagnosis. In such cases, the majority classes tend to bias the classifiers to one another, and the classifier represents the minority class classification rates incorrectly. Finally, the classifier analyses the majority class entirely and the minority class was excluded.^(3,4,5,6)

Numerous methods were offered in research to address issues related to the CI. This is a demanding and stimulating study issue in a period where numerous concerns require attention simultaneously, including the size of imbalanced datasets, cost of misclassified class, class overlapping, MC issue, BC issue, and minor disjoints. While (BCI) Binary CI problems involving imbalanced data have been addressed, (MCI) Multi CI problems involving a variety of issues categories are rarely resolved. wherein there are several CI issues where there may be one or more majority and minority classes. It is still necessary for providing consideration for MC problems even though decomposition or other approaches may be employed. So, it will be extremely daring to move forward with the minority class whenever the data are divergent in nature.⁽⁷⁾

Because of the significance of this topic, considerable efforts have been made to determine methods to address these challenges. These assertions can be distributed into 3 types according to the way approaches address the issue of CI: external approaches, or data level approaches, which involve preparing data to balance class distributions and reduce the impact of differential distribution during the process of classification. When creating new algorithms or altering current ones, the internal or algorithmic level technique considers the effects of minor class differences. The third strategy is the cost-sensitive technique, which has the potential to combine algorithmic and data-level techniques to include diverse misclassification rates for every class throughout the learning stage. Pre-processing and algorithmic methods are very effective among them.

To avoid these issues in recent work proposed map reduce based data parallelization of class imbalanced datasets. A new Over Sampling (OS) technique called MOKCCAS has been suggested for minimizing data loss throughout FSP. This technique takes advantage of the constant FE ability to utilize a variant of the ASSOMs that is derived from KCCA. Nevertheless, the noisy data values in the dataset employed for this study could have an impact on classification and learning outcomes.

To overcome those issues this, work first introducing a preprocessing using binning to remove unwanted noise from the input. And then data parallelization will be done by using map reduce framework to solve this computation requirement problem. And then reducing data loss throughout FSP, this work suggests a new oversampling procedure, named MOKCCAS. For simulating various aspects of the input data distribution from which the synthetic cases produced by various subspaces will receive, many subspaces are utilized from the

parallel phase. Then proposes a feature selection model using Mutated whale optimization (MWO) methods and produces more significant features and reduces the time consumption. Finally proposed class balancing model will be tested using uniform distribution based enhanced adaptive neuro fuzzy inference system (UDANFIS).

Related Works

A novel Under Sampling (US) technique was presented by Vuttipittayamongkol *et al.*⁽⁸⁾; it removes undesirable occurrences from the area of overlapping and increases the prominence of the minority occurrences as a result. The suggested approach demonstrated statistically substantial increases in classification performance when tested and evaluated on 36 publicly available imbalanced datasets.

A Weighted Kernel-based SMOTE (WK-SMOTE) has been suggested by Mathew *et al.*⁽⁹⁾ for overcoming SMOTE's drawback for nonlinear situations by OS the Support Vector Machine (SVM) classifier's feature space. Using many benchmarking unbalanced data sets, it is demonstrated that the suggested OS methodology combined with a cost-sensitive SVM framework improves performance over alternative standard approaches. Furthermore, a hierarchy-based system with a progressive class order is designed for MCI issues. For the purpose of detecting deprivation in the separation of high-voltage equipment, the recommended WK-SMOTE and hierarchical architecture are verified on an actual industrial (FD) Fault Detection situation.

A Maximum Margin of Twin Spheres SVM (MMTSSVM) has been suggested by Xu⁽¹⁰⁾ for the categorization of imbalanced information. Only two homocentric spheres must be located via MMTSSVM. The large sphere drives most samples into the minority class by widening the gap among the 2 homocentric spheres, whereas the small sphere attempts to identify as many samples as possible in the majority class. Due to the utilization of QPP and LP (Linear Programming), the computational speed of MMTSSVM is significantly increased, as compared to two QPPs in standard TSVM or a larger QPP in SVM. The matrix inverse operation is avoided by MMTSSVM, which is more significant. One simulated study is employed to analyze and verify the MMTSSVM's parameter feature.

As contrast with State of The Art (SOTA) approaches, experimental outcomes on 9 benchmark datasets indicate the efficacy of the suggested MMTSSVM is, at last, utilize MMTSSVM to improve the experimental outcome of a clinical study on Alzheimer's disease.

Multiclass radial-based oversampling (MC-RBO), a new DS (Data-Sampling) technique specifically designed for MC situations, has been suggested by Krawczyk *et al.*⁽¹¹⁾. Utilizing potential functions to create unreal instances represents the method's key innovation. Considering data from all classes is important, as opposed to current MCO (Multiclass Oversampling) methods that primarily employ features of minority classes. The exploration of regions where the mutual class distribution has limited value serves as a guide for the false instance generation procedure. Ensure an innovative oversampling process in this manner, which can address the drawbacks of current techniques and handle challenging data distributions. A comprehensive statistical analysis and a comprehensive experimental investigation are employed to assess the utility of the MC-RBO technique. The outcomes demonstrate the ability to learn MCI data may be greatly enhanced by accounting for data from every class while performing effective oversampling.

Active Online-Weighted ELM (AOW-ELM), a very effective solution depends on the Extreme Learning Machine (ELM) classification framework, is presented by Yu *et al.*⁽¹²⁾. This study's primary benefits are as follows: 1) Weighted ELM (WELM) is chosen as the base classifier to ensure the objectivity of case selection in the active learning way. 2) For selecting originally designated instances for minimizing the undetected cluster influence and cold start issue as effectively as probable via the hierarchical clustering technique; and 3) a well-organized online updated means of WELM is theoretically assumed. These are the reasons regarding an imbalanced instance distribution can disrupt (AL) Active Learning. 4) More adaptable early stopping criteria is suggested, which is comparable to the margin exhaustion condition. The suggested AOW-ELM technique is more effective than other SOTA AL techniques that are especially built for the CI situation, as demonstrated by experiments conducted on 32 BC data sets with varied imbalance ratios.

In this type of MC setting, the SCUT HS (Hybrid Sampling) technique, introduced by Agrawal *et al.*⁽¹³⁾, is utilized to balance the quantity of training samples. By creating synthetic cases, the SCUT method oversamples samples from minority classes while US majority classes using cluster analysis. Furthermore, it manages imbalances among and between classes. Testing the SCUT approach for pre-processing the data prior to classification delivers extremely accurate models that outperform the SOTA, according to experimental outcomes instead of several MC issues.

A scaling kernel-SVM method was presented by Zhang *et al.*⁽¹⁴⁾ to address the MCI data classification issue. To obtain an approximate hyperplane, first apply the conventional SVM method. The scaling kernel function should then be presented, and its parameters should be determined through the Weighting Factors (WF) and Chi-Square Test (CST). The suggested approach has robust adaptability and can address the classifier performance deterioration issue caused by Skewed Data Distribution (SDD), according to experimental outcomes on KEEL data sets.

METHOD

This section stages the proposed imbalanced data handling framework, that has 7 stages initially it preprocess the input data using Binning to smooth the data, second one is data Parallelism using map reducer framework, third one is Minority Oversampling using Kernel Canonical Correlation Analysis subspace, fourth one is Competitive learning for finding the basis vector ,fifth one is Orthonormalization based normalization, sixth one is Features Selection (FS) using Mutated whale optimization and seventh one is classification using uniform distribution based enhanced ANFIS. ^(15,16,17,18,19,20) Figure 1 presents the recommended system’s common structure.

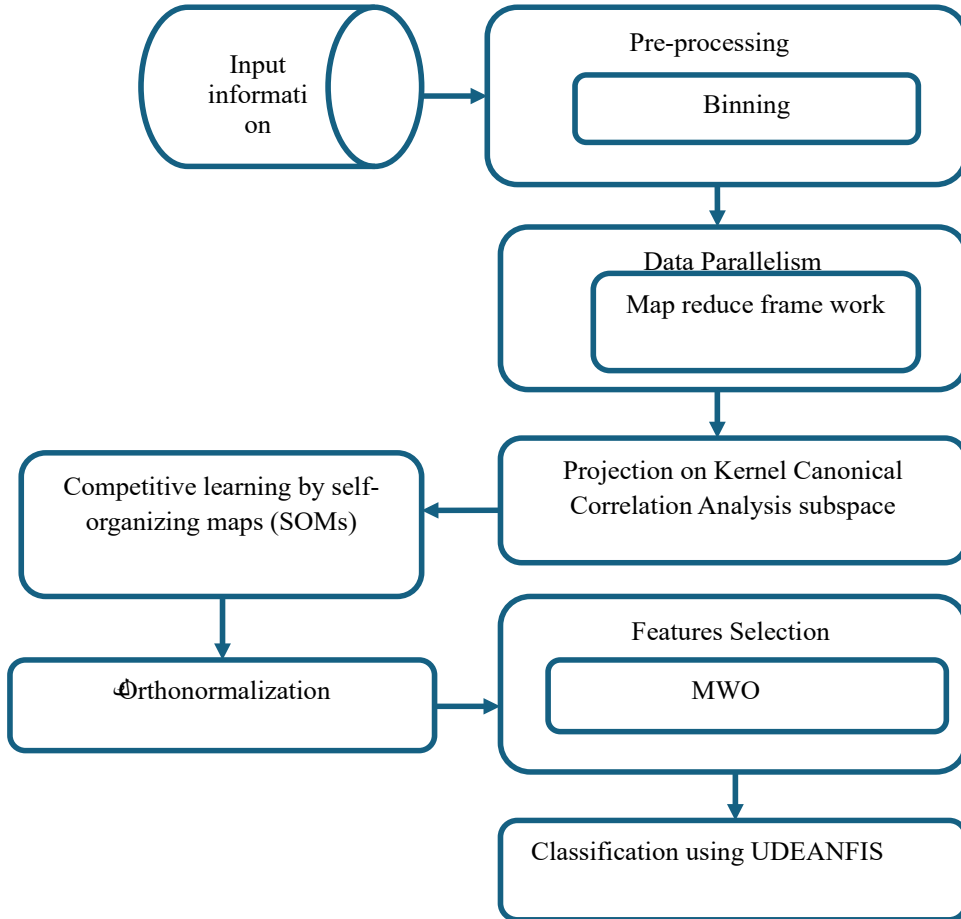


Figure 1. Overall structural design of the suggested framework

Algorithm for Mutated Whale Optimization

START

1. import information
2. set the locations of the whale population X
3. calculate the fitness of every whale
4. set a and r, calculate A and C
5. set X* as the best hunter whale’s position
6. set X** as the best solution from mutation
7. set t = 1
8. while t ≤ max iterations do
9. for every hunting whale do
10. if p < 0,5
11. if |A| < 1
12. update the present hunting whale’s position with (7)
13. else if |A| ≥ 1
14. arbitrarily choosing another search agent
15. update the present hunting whale’s position with (9)
16. end if

17. else if $p \geq 0,5$
18. update the present hunting whale's position with (8)
19. end if
20. end for
21. update X^* if there is a better solution
22. Apply binary mutation operator
23. Get new set of solutions
24. calculate the fitness of every whale
25. update X^{**} if there is a better solution
26. $t = t + 1$
27. end while
28. output X^{**}

END

Classification Using Uniform Distribution Based Enhanced ANFIS

After feature selection the selected features are need to classify to test the proposed minority oversampling model. This work uses Uniform Distribution Based Enhanced ANFIS for classification. Situated on the Neuro Fuzzy Network (NFN), the ANFIS network is one of the NN (Neural Network) types. Figure 2 represents the ANFIS structure utilized for executing these 2 rules.^(21,22,23,24,25)

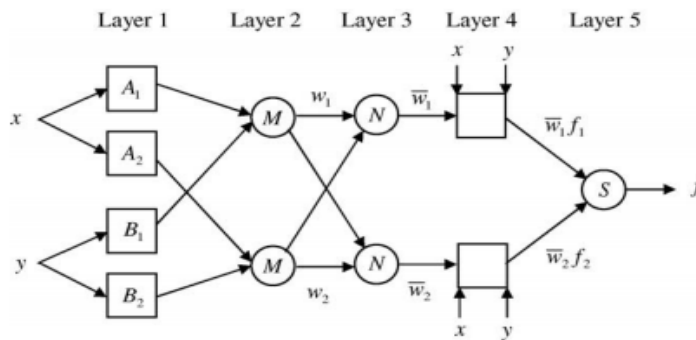


Figure 2. ANFIS architecture

Uniform Distribution (UD) Based Enhanced ANFIS

In traditional ANFIS initial layer uses gauss ion distribution to compute the Membership Function (MF) for data samples. The negative infinity of the ND (Normal Distribution) at the beginning is one of its drawbacks when it comes to reliability analyses. Certain outcomes may have negative values as an outcome of this.

For overcoming those problems this study utilizes UD which starts from zero for membership value computation in first layer of ANFIS which is known as Uniform Distribution Based Enhanced ANFIS.

All of the nodes in the initial layer are adaptable. Layer 1's outputs are the inputs' Fuzzy Membership (FM) levels, that can be obtained by:

$$o^2 = u_i^{(j)} (a_i^{(2)}) \quad (1)$$

Here: $u_i^{(j)} (.) : R \rightarrow [0,1], (i=1,2,\dots,M, j =1,2,\dots,N)$, and MF can be represented as $u_i^{(j)} (.)$

$$\mu (x) = \frac{1}{1 + (\frac{x-c_i}{a_i})^{2b_i}} \quad (2)$$

The premise parameters sets are denoted as $a_i, b_i,$ and c_i .

Fixed nodes can be identified in the 2nd layer. They function as a fundamental multiplier, as indicated by the letter M on its label. The following is a representation of this layer's outputs:

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y), i = 1,2 \quad (3)$$

Here, (x) and (y) can be represented as MFs, and the linguistic labels can be denoted as A and B . A firing strength of rule is indicated by the output w_i . It is indicated by each node's outcomes.

Fixed nodes are also present in the 3rd layer. Normalization role for firing strengths from the previous layer can be labelled by N. The following represents this layer’s outputs:

$$O_i^3 = w_i = \frac{w_1}{w_1+w_2}, i = 1,2 \quad (4)$$

The normalized firing strengths represent the layer’s outputs.

Adaptive nodes constitute the nodes in the 4th layer. Every node in this layer produces an output that is simply obtained by the normalized firing strength’s product. Consequently, this layer’s outputs have been supplied through:

$$O_i^4 = w_i f_i = w_i(p_i x + q_i y + r_i) \quad (5)$$

Here, the parameter set is indicated by $\{p_i, q_i, r_i\}$, and 3rd layer output can be denoted as O_i^3 . The term “resulting parameters” refers to those parameters.

There exists only one fixed node, designated S, in the 5th layer. Then adding all incoming signals is done by this node. Therefore, the model’s overall output can be obtained via:

$$O_i^5 = \sum_{i=1} w_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (6)$$

RESULTS AND DISCUSSION

This part analyses the recommended algorithm’s simulation results. The framework is implemented using MATLAB. The Radial Based Oversampling (RBO), Ensemble based adaptive over-sampling algorithm (EBAOA), Minority Oversampling in Kernel Canonical Correlation Analysis (KCCA) Subspaces and proposed Kernel canonical correlation Analysis subspace with Uniform distribution based enhanced Adaptive Neuro Fuzzy Inference System (KA-UANFIS) are compared by the precision, recall, Error Rate (ER), then from the UCI (ML) Machine Learning repository, Accuracy for the dataset approaches on 10 real-world benchmark imbalanced datasets.

Table 1. Performance Comparison Results

Metrics	Method			
	RBO	EBAOA	KCCA	KA-UANFIS
Accuracy	86	89,063	91,797	95,443
Precision	82,51	87,661	90,451	96,058
Recall	86,29	88,743	92,228	95,461
Error Rate	14	10,938	8,2031	4,5573

Precision

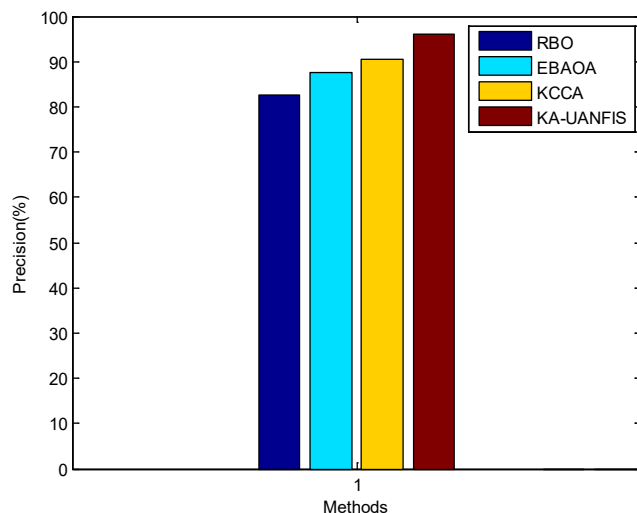


Figure 3. Precision

The comparison metric is assessed by the precision, as can be observed in figure 3. The procedures are represented on the x-axis, and on the y-axis, the precision outcomes are plotted. With assumed dataset, the approaches such as existing RBO, EBAOA and KCCA procedures offer low precision results which is 82,51 %, 87,66 % and 90,45 % whereas proposed KA-UANFIS model provides higher precision results which is 96,05 %. So, the outcome indicates that the KA-UANFIS upsurges the precision of classification via more informative structures which are selected in this work using mutated whale optimization algorithm which uses fitness function evaluation to select features by this true positive rate is improved.

Recall

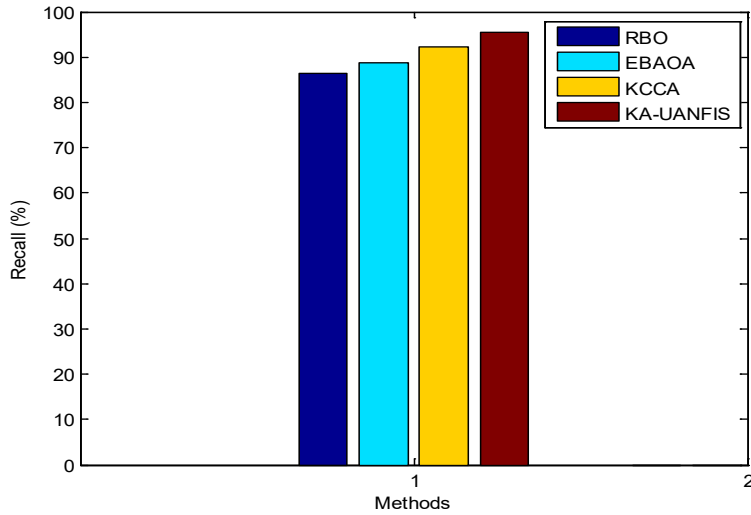


Figure 4. Recall

The recall performance comparison outcomes for the RBO, EBAOA, and KCCA techniques already in practice and the suggested KA-UANFIS are presented in figure 4. The various classification techniques are marked on x-axis, and the y-axis shows the recall value. The suggested study FS using the WO framework, in which the global solution is calculated using the binary mutation function. This raises the true positive rate, which in turn raises the recall automatically. The aforementioned figure indicates that, for the given dataset, the suggested KA-UANFIS technique yields a higher recall result of 95,46 %, while the RBO, EBAOA, and KCCA approaches techniques currently in practice yield lower recall outcomes of 86,29 %, 88,74 %, and 92,22 %.

Error rate

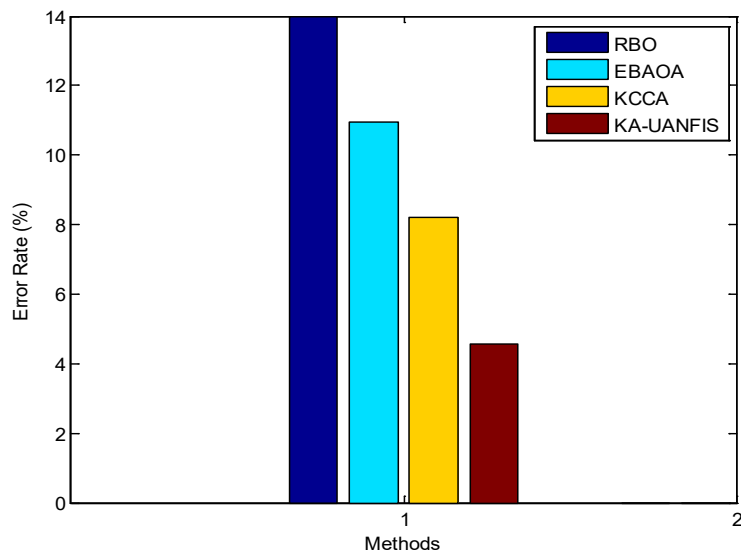


Figure 5. Error rate

Performance comparison results in terms of error rate is shown in the above figure for the proposed KA-UANFIS and existing RBO, EBAOA and KCCA methods. In the above graph x-axis represents different classification methods and y-axis represents the error rate value. From the figure 5 it is determined that the suggested KA-UANFIS algorithm provides lower error rate result which is 4,55 % while the current RBO, EBAOA and KCCA methods provides higher Error rate results which are 14 % ,10,93 % and 8,2 % for the given dataset.

Accuracy

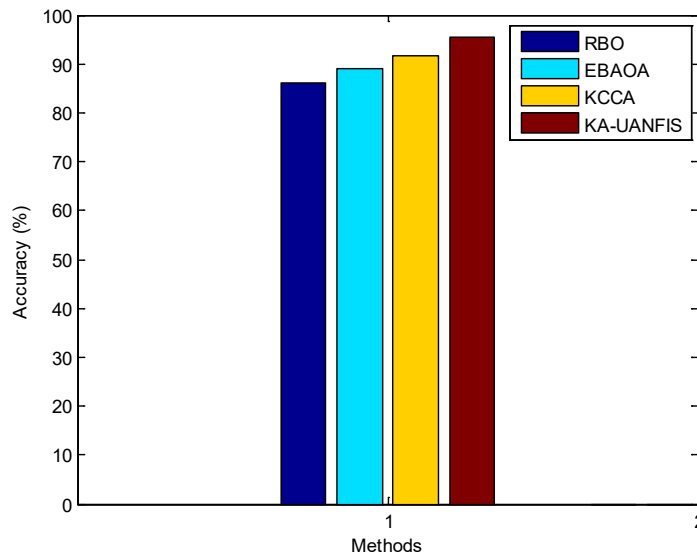


Figure 6. Accuracy

Figure 6 presents a comparison of accuracy metrics for several classifiers, including KA-UANFIS, KCCA, EBAOA, and RBO, where the accuracy value is represented on the y-axis and the techniques are plotted on the x-axis. From the above figure it is concluded that the proposed KA-UANFIS algorithm provides higher accuracy result which is 95,44 % whereas the existing RBO, EBAOA, KCCA methods provides lower accuracy outcomes as 86 %, 89,06 % and 91,79 % for the given gene expression dataset.

CONCLUSIONS

This work aimed to suggest an improved framework for imbalanced data handling. In which first introducing a preprocessing using binning to remove unwanted noise from the input. And then data parallelization will be done by using map reduce framework to solve this computation requirement problem. The paper then suggests a new oversampling approach called MOKCCAS that reduces data loss throughout FSP. For simulating various aspects of the input data distribution from which the synthetic instances produced by various subspaces will acquire when many subspaces are utilized from the parallel phase. Then proposes a feature selection model using Mutated whale optimization (MWO) methods and produces more significant features and reduces the time consumption. Finally proposed class balancing model will be tested using uniform distribution based enhanced adaptive neuro fuzzy inference system (UDANFIS). Experimental results shows that this proposed model delivers superior outcomes than other all SOTA approaches in terms of precision, recall and accuracy. Future work planned to use hybrid feature selection model to enhance the accuracy.

BIBLIOGRAPHIC REFERENCES

1. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, and Bing G. Learning from class-imbalanced data: Review of methods and applications. *Expert systems with applications*, 73, pp. 220-239. <https://doi.org/10.1016/j.eswa.2016.12.035>.
2. Gosain A, and Sardana S. Handling class imbalance problem using oversampling techniques: A review. In *international conference on advances in computing, communications and informatics (ICACCI)*, pp. 79-85. <https://doi.org/10.1109/ICACCI.2017.8125820>.
3. Jian C, Gao J, and Ao Y. A new sampling method for classifying imbalanced data based on support vector machine ensemble. *Neurocomputing*, 193, pp. 115-122. <https://doi.org/10.1016/j.neucom.2016.02.006>.

4. Menardi G, and Torelli N. Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery*, 28, pp. 92-122. <https://doi.org/10.1007/s10618-012-0295-5>.
5. Błaszczyński J, and Stefanowski J. Neighbourhood sampling in bagging for imbalanced data. *Neurocomputing*, 150, pp. 529-542. <https://doi.org/10.1016/j.neucom.2014.07.064>.
6. Oksuz K, Cam BC, Kalkan S, and Akbas E. Imbalance problems in object detection: A review. *IEEE transactions on pattern analysis and machine intelligence*, 43(10), pp. 3388-3415. <https://doi.org/10.1109/TPAMI.2020.2981890>.
7. Thabtah F, Hammoud S, Kamalov F, and Gonsalves A. Data imbalance in classification: Experimental evaluation. *Information Sciences*, 513, pp. 429-441. <https://doi.org/10.1016/j.ins.2019.11.004>.
8. Vuttipittayamongkol P, Elyan E, Petrovski A, and Jayne C. Overlap-based undersampling for improving imbalanced data classification. In *Intelligent Data Engineering and Automated Learning-IDEAL: 19th International Conference, Proceedings, Part I* 19 pp. 689-697. https://doi.org/10.1007/978-3-030-03493-1_72.
9. Mathew J, Pang CK, Luo M, and Leong WH. Classification of imbalanced data by oversampling in kernel space of support vector machines. *IEEE transactions on neural networks and learning systems*, 29(9), pp. 4065-4076. <https://doi.org/10.1109/TNNLS.2017.2751612>.
10. Xu Y. Maximum margin of twin spheres support vector machine for imbalanced data classification. *IEEE transactions on cybernetics*, 47(6), pp. 1540-1550. <https://doi.org/10.1109/TCYB.2016.2551735>.
11. Krawczyk B, Koziarski M, and Woźniak M. Radial-based oversampling for multiclass imbalanced data classification. *IEEE transactions on neural networks and learning systems*, 31(8), pp. 2818-2831. <https://doi.org/10.1109/TNNLS.2019.2913673>.
12. Yu H, Yang X, Zheng S, and Sun C. Active learning from imbalanced data: A solution of online weighted extreme learning machine. *IEEE transactions on neural networks and learning systems*, 30(4), pp. 1088-1103. <https://doi.org/10.1109/TNNLS.2018.2855446>.
13. Agrawal A, Viktor HL, and Paquet E. SCUT: Multi-class imbalanced data classification using SMOTE and cluster-based undersampling. In *7th international joint conference on knowledge discovery, knowledge engineering and knowledge management (IC3k)*, 1, pp. 226-234.
14. Zhang Y, Fu P, Liu W, and Chen G. Imbalanced data classification based on scaling kernel-based support vector machine. *Neural Computing and Applications*, 25, pp. 927-935. <https://doi.org/10.1007/s00521-014-1584-2>.
15. Ahmed MM, Houssein EH, Hassanien AE, Taha A, and Hassanien E. Maximizing lifetime of wireless sensor networks based on whale optimization algorithm. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, pp. 724-733. https://doi.org/10.1007/978-3-319-64861-3_68.
16. Kumawat IR, Nanda SJ, and Maddila RK. Multi-objective whale optimization. In *Tencon ieee region 10 conference*, pp. 2747-2752. <https://doi.org/10.1109/TENCON.2017.8228329>.
17. Sharawi M, Zawbaa HM, and Emary E. Feature selection approach based on whale optimization algorithm. In *Ninth international conference on advanced computational intelligence (ICACI)*, pp. 163-168. <https://doi.org/10.1109/ICACI.2017.7974502>.
18. Salgotra R, Singh U, and Saha S. On some improved versions of whale optimization algorithm. *Arabian Journal for Science and Engineering*, 44, pp. 9653-9691. <https://doi.org/10.1007/s13369-019-04016-0>.
19. Ling Y, Zhou Y, and Luo Q. Lévy flight trajectory-based whale optimization algorithm for global optimization. *IEEE access*, 5, pp. 6168-6186. <https://doi.org/10.1109/ACCESS.2017.2695498>.
20. Qiao W, Huang K, Azimi M, and Han S. A novel hybrid prediction model for hourly gas consumption in supply side based on improved whale optimization algorithm and relevance vector machine. *IEEE access*, 7, pp. 88218-88230. <https://doi.org/10.1109/ACCESS.2019.2918156>.

21. Janalipour M, and Mohammadzadeh A. Building damage detection using object-based image analysis and ANFIS from high-resolution image (case study: BAM earthquake, Iran). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), pp. 1937-1945. <https://doi.org/10.1109/JSTARS.2015.2458582>.

22. Rubio JDJ, Cruz DR, Elias I, Ochoa G, Balcazar R, and Aguilar A. ANFIS system for classification of brain signals. *Journal of Intelligent & Fuzzy Systems*, 37(3), pp. 4033-4041. <https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs190207#:~:text=DOI%3A-,10.3233/JIFS%2D190207,-Journal%3A%20Journal>.

23. Mlakić D, Baghaee HR, and Nikolovski S. A novel ANFIS-based islanding detection for inverter-interfaced microgrids. *IEEE Transactions on Smart Grid*, 10(4), pp. 4411-4424.

24. Selvapandian A, and Manivannan K. Fusion based glioma brain tumor detection and segmentation using ANFIS classification. *Computer methods and programs in biomedicine*, 166, pp. 33-38. <https://doi.org/10.1016/j.cmpb.2018.09.006>.

25. Priyadarshini L, and Shrinivasan L. Design of an ANFIS based decision support system for diabetes diagnosis. In *International Conference on Communication and Signal Processing (ICCSP)*, pp. 1486-1489. <https://doi.org/10.1109/ICCSP48568.2020.9182163>.

FINANCING

The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Krishnakumar V.

Data curation: Sangeetha V.

Formal analysis: Sangeetha V.

Research: Krishnakumar V.

Methodology: Krishnakumar V.

Drafting - original draft: Sangeetha V.

Writing - proofreading and editing: Krishnakumar V.