ELM-Based Imbalanced Data Classification-A Review

*[1]Brajendra Singh Rajput, [1]Partha Roy, [1]Sunita Soni, [2]Bhagat Singh Raghuwanshi
[1]Computer Science and Engineering, Bhilai Institute of Technology, Durg, Chhattisgarh, India, Chhattisgarh Swami Vivekanand Technical University (CSVTU)
[2]Information Technology, Madhav Institute of Technology and Science (MITS) Gwalior, Madhya Pradesh, India
E-mail: brajendrarajput89@gmail.com, partha.roy@bitdurg.ac.in, sunita.soni@bitdurg.ac.in, bhagat@mitsgwalior.in
*Corresponding Author

Keywords: SMOTE, Imbalanced Learning, Classification, under bagging, ensemble, Voting approach, variable length, Kernelized ELM, Fraud prediction.

***Recieved**:* August 8, 2023

Imbalance issues occur in Machine Learning (ML) when there is high distortion in the class distributions. A great challenging task in ML is the imbalance of data classification. It is because most classification methodologies tend to bias toward the majority class even though high importance is given to the minority class. To enable its stable operation, many techniques are utilized recently that are still in use for classifying imbalanced datasets efficiently. Owing to the assumption with balanced class distribution or equal misclassification, the prevailing learning algorithms are prone to favor the majority class when handling complicated classification issues with skewed class distribution. The most prominently adopted technique to deal with data having imbalance class distribution is Extreme Learning Machine (ELM). Unwanted class boundaries as of data with unbalanced classes may be learned by ELM similar to various other classification algorithms. Grounded on the kernel utilized, elevated weighted ELM, active learning-centered techniques, etc, an augmentation in the ELM framework is done for efficient imbalanced classification. Regarding ELM approaches, the latest studies for imbalance classification are studied here. Finally, regarding G-Mean, Accuracy, and Imbalance Ratio (IR), the research studies' performance was analogized.

Povzetek: Prispevek obravnava neuravnotežene učne primere. Podaja pregled zlasti novejših metod za klasifikacijo neuravnoteženih podatkov z ELM in poudarja izzive pristranskosti modelov do večinskih razredov ter opisuje tehnike, kot so SMOTE in jedrski ELM.

1 Introduction

Data stream classification has garnered wide interest in the modern era owing to the massive expansion in data availability on the Internet and in other fields. Unstructured data streams that consistently arrive on time are difficult to classify because they lack class labels along with accumulating over time [1]. Many data stream algorithms do not perform satisfactorily or fail altogether in mixed data streams comprising categorical along with numerical values or in limited labeled samples. Multi-classes have more categories than two, while binary classes contain only two categories in the dataset [2].

The dataset is affected by imbalance problems like text classification, web fault prediction, Credit Card Fraud Detection (CCFD), high error rate classification models, etc. [3]. The time-dependent alteration in data streams is one major issue in data stream classification schemes [4]. Another typical classification issue is Class Imbalance (CI) which develops when one of the classes, known as the minority class, has a smaller amount than another class, known as the majority class. To deal with unstructured class streams, many strategies were engendered. Unsupervised learning and supervised learning are the two diverse learning strategies that are employed to address CI issues in ML paradigms [5]. Resampling, cost-sensitive learning, one-class learning, feature selection, and other techniques are also used. However, multiclass classification was not accurately dealt with by conventional methodologies [6]. Hence, the imbalance issue engendered in the data classification process was solved by ELM. Figure 1 shows a diagrammatic representation of imbalance classification.

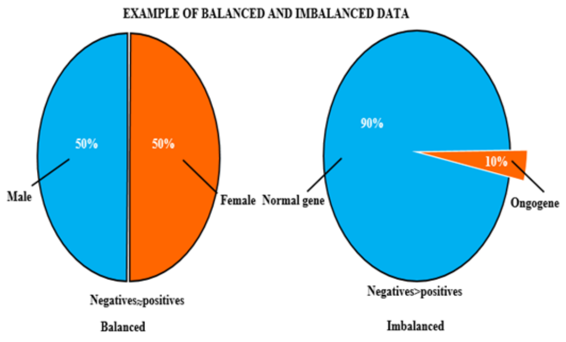

Figure 1: Diagrammatic representation of imbalance classification.

A Single Hidden Layer Feed-Forward Neural Network (SHLFN) is the foundation of an ELM. Utilizing kernels, random neurons with or without specified shapes, and optimization constraints are all combined in this method [7]. To determine the SHLFN's output weights, it randomly chooses the hidden units [8]. Owing to the lack of iterative tuning with training utilizing the generalization operation, ELM is quicker than SHLFN. Additionally, the fundamental benefit of employing ELM is that it models a standard approach for both binary class and multi-class situations [9] in addition to offering universal approximation capability and classification ability. Medical data classification, medical diagnostics, image quality evaluation, and many more are the several application disciplines of ELM. Moore-Penrose generalized inverse is utilized by ELM to update its weight apart from standard single hidden neural networks. The classification accuracy is maximized by several prevailing extreme learning algorithms by classifying the majority class accurately while incorrectly classifying the minority class [10]. For classifying imbalanced datasets, a variety of sampling-based strategies, including under-sampling and oversampling, cost-sensitive learning techniques, along with ensemble learning, have recently been employed

2 Related works

Grounded on the ELM approach, diverse prevailing research methodologies related to the imbalanced classification problem were explained as follows,

2.1 ELM algorithm-based approaches for imbalanced class problem

Bhagat *et al.* [11] presented a novel Synthetic Minority Oversampling Technique centered Class-Specific ELM (SMOTE-CSELM) for elevating imbalanced classification complexities with the skewed class distribution. The minority class samples were elevated by SMOTE that created synthetic samples pertained to the minority class to determine the classifier's decision region. Regarding computation complexity, the experiential outcomes illustrated the presented approach's higher efficacy. However, the large size imbalanced classification complexity could not be tackled by the presented model.

Bhagat Singh *et al.* [12] developed Under Bagging ensemble-centered variants of a Kernelized ELM (UBKELM) where the random under-sampling's strength along with bagging was integrated. By a majority of class samples' random under-sampling, balanced training subsets of large numbers were integrated. As UBKELM had more stability along with promising generalization performance, it was utilized as a component classifier to execute ensemble operations. However, the approach's accuracy was impacted by the dependence of several training subsets on the CIs degrees.

Sanyam Shukla *et al.* [13] explicated a CS-ELM's variant named Online Sequential Class-Specific ELM (OSC-SELM). Here, it utilized a class-specific regularization approach that makes assigning weights to the training samples unnecessary. The class-specific regularization coefficients were computed by utilizing the class proportion along with the regularization coefficient's values. Thus, better generalization was attained by the presented scheme than various other prevailing schemes. However, the presented method's stability was affected by the overfitting complexities.

Hualong *et al.* [14] projected a proficient Active Online-Weighted ELM (AOW-ELM) classification scheme to tackle larger time consumption issues. To tackle the CI problem, a cost-sensitive learning scheme was adopted to choose the Weighted ELM (WELM). An active learning model was built by utilizing an AL-ELM algorithm, subsequently. For an effective weight update rule, an efficient online learning mode of WELM was designed. Finally, a flexible along with effective early stopping criterion was deduced centered on the margin exhaustion criterion. The missed clusters were avoided by the utilized clustering techniques. By this, the suggested method was recognized to be more efficient than numerous prevailing models. However, the output was affected by the random initialization of weights variation betwixt the input and HL.

Wendong et al. [15] suggested a scheme for classification issues with imbalanced data distributions by utilizing the Class-specific Cost Regulation ELM (CCR-ELM) algorithm. Kernel-centered CCR-ELM was formed by introducing kernel functions in CCR-ELM. The number of class sample's effects was pondered in CCR-ELM along with the effects of data's dispersion degree. An elevation in classification performance with the superior status diagnosis was explicated in the experiential outcomes. However, direct utilization in multiclass imbalanced classification issues was not handled by the presented model.

Bhagat *et al.* [16] provided a variant of ELM named Class-Specific ELM (CS-ELM) for handling binary CI issues. For better-imbalanced classification, class-specific regularization parameters were employed here, that are computed utilizing class distribution. As there was no weight assigning the process to the training instances, the presented model's performance was elevated. However, the ELM's stability cannot be addressed by the presented scheme under several operational uncertainties.

Yong *et al.* [17] developed a Differential Evolution (DE)-centered ensemble learning stratagem to resolve the imbalanced data classification problem. Numerous WELMs were chosen initially with distinct activation functions as base learners to embody the university. Additionally, each base learner's weight was optimized by utilizing the DE algorithm. A candidate weight vector's population was presented subsequently. Then, choosing the individual having best fitness value as the base learner's weight in the ensemble occurs. Regarding Geometric mean (G-mean), a superior classification performance than other analogized models was provided by the suggested approach. Nevertheless, the technique’s performance was affected by the outlier’s attendance.

Hui *et al.* [18] explained an Evolutionary ELM with a sparse cost matrix for imbalanced learning. Here, engendering the case-weighting ELM on a sparse cost matrix occurred in a diagonal form. The misclassification issue was optimized by the multi-objective optimization regarding penalty factors utilizing an evolutionary algorithm merged with an error-bound model. As the presented model was aided by the link betwixt the generalization ability and case-weighting factors, it was utilized as adaptive cost-sensitive learning. However, identifying the penalty factor’s perfect set along with thumb rules was critical as it was too specific to fit generic situations.

Chengbo *et al.* [19] introduced an Improved Weighted-ELM (IW-ELM) algorithm for imbalanced classification. The appropriate weights were assigned for the imbalanced data classification by presenting a voting-centered weighting scheme. The weighted ELM classifier’s training was involved in the presented approach. Following that, proper classifiers for voting were determined by eliminating unusable classifiers. And finally, grounded on majority voting, the classification outcomes were determined. When analogized with other ELM-centered algorithms, high accuracy was exhibited by the simulation outcomes. But, the training time was elevated by the difficulty in the misclassification cost matrix’s generation

Tianlei *et al.* [20] employed a Deep-Weighted ELM (DWELM) algorithm for imbalanced data classification. The representation capacity was augmented by the elevated stacked multilayer deep representation network that was trained with ELM (EH-DrELM). The sample weights for imbalanced multiclass data were optimized by a rapid AdaBoost algorithm. The data imbalance issue in sequential learning was alleviated by meta-cognitive online sequential ELM. However, pushing the data as of distinct classes in diverse directions was not sufficiently made by the linear ELM.

Honghao *et al.* [21] established Weighted ELM (WELM) to optimize imbalanced classification. Grounded on the dandelion's behavior, a new swarm intelligence algorithm named Dandelion Algorithm (DA) was employed for imbalance classification. A diverse number of seeds are engendered by every dandelion in DA. The next iteration's dandelion was opted as of the seeds formed by diverse dandelions. The dandelion's position was altered via sowing along with selection operations. High detection performance was exhibited by the experiential outcomes than other conventional schemes. It was also utilized in CCFD. However, the system's efficacy was affected by the degradation in convergence rate.

2.2 Ensemble-based ELM imbalanced classification problem.

In table1, CI problem classification based on ensemble-based ELM is explained.

Table 1: Analysis of Ensemble-based ELM imbalanced classification problem

Researcher Name	Methodology	Description	Result	Drawback
Arnis *et al.*[22]	Entropy-centered Classifier (EC) algorithm.	Grounded on the original class proportions in the training dataset, weights were encompassed in the entropy computation initially. For each class, the weights or class importance was computed that remains unaltered during the learning process.	Regarding classification accuracy along with sensitivity, a more promising solution was attained in a complex environment.	If the initial dataset was unbalanced, a high false negative rate was shown.
Xiaoyan *et al.*[23]	Ensemble of ML- K-Nearest Neighbour (EML-KNN)	Metadata extraction was the initial step that encompasses meta-target identification along with the meta-feature collection. In the model construction stage, a meta-model was then constructed for each candidate's algorithms. Finally, a new classification problem was recommended in the recommendation step.	Augmented recommendation performance was resulted by recommending diverse algorithms for various classification issues automatically.	The improper distance method affected the quality.
Wentao *et al.*[24]	Sparse Weighting centered on Online Sequential ELM (SW-OSELM)	By oversampling utilizing the SMOTE, a balanced training set was attained initially. Grounded on training errors, a balanced process was executed after that. Centered on the final training set, the initial mode was set.	Better classification accuracy with diminished accuracy loss was exhibited by the experiential outcomes.	Unnecessary training time results in the presence of redundant virtual samples.
Junhai *et al.*[25]	Map Reduce and	Utilizing the Map Reduce	Classifying the imbalance	The system's

	Ensemble-centered ELM technique.	algorithm, the positive class instance's center was computed initially. The sampling was executed followed by the classifier component's training utilizing the ELM algorithm. By utilizing the voting approach, the integration was executed finally.	d large datasets elevated the positive class instance's learning region.	performance was affected by the big data partitioning into small pieces automatically along with their deployment in a parallel computing node.
Shifei *et al.*[26]	UnSupervised-ELM(US-ELM)	By framing a new cost function, the embedded matrix was obtained initially. The clustering in the embedded matrix was executed by the k-means algorithm. Thus, input weights along with biases are generated randomly. Further, the original data's main features are obtained by ELM as an autoencoder (ELM-AE).	faster responding along with greater generalization	It was not suitable for high-dimensional datasets as it takes a longer training time.

Fulong *et al.* [27] presented an ensemble-centered adaptive over-sampling scheme to overcome the CI problem. Elevated microaneurysm detection was attained by utilizing Boosting and Bagging along with Random subspace. Thus, the positivity of adaptive over-sampling and ensemble was integrated by the presented ensemble-centered over-sampling schemes. The induction biases introduced as of the imbalanced data were diminished by the amalgamation of ensemble and adaptive over-sampling. Generalization along with ELM classification performance was elevated as a result of this. However, the classification of imbalanced data was made more critical owing to the presence of false alarms.

Zhenyu *et al.* [28] developed the Easy-SMT ensemble approach to tackle the imbalance of learning's impact. With a SMOTE-centered oversampling policy to supplement minority defective classes along with Easy Ensemble to alter a CI learning problem into an ensemble-grounded balanced learning subproblem, Easy-SMT was nothing more than an integrated ensemble-based approach. Regarding good classification capability, the presented method outperformed other conventional methods. Because of the minority class sample's small-sized features, a large number of minority class samples were not recognized clearly.

Guillem *et al.* [29] projected a Probability Threshold (PT) - Bagging approach to solving CI problems raised in the network. Cost-sensitive learning, rebalancing mechanisms, along with threshold moving were the '3' stages in the presented PT-Bagging approach. Various misclassification costs were assigned to different classes initially. To balance the training data, the data was resampled after that. By utilizing the corresponding threshold, a model as of the dataset was altered into a class label. As the presented work discarded various training data, more efficient computation than the traditional schemes was attained by it. However, the error rate was elevated by the miscalibrated posterior probability and the prior shift. Shown in Figure 2.

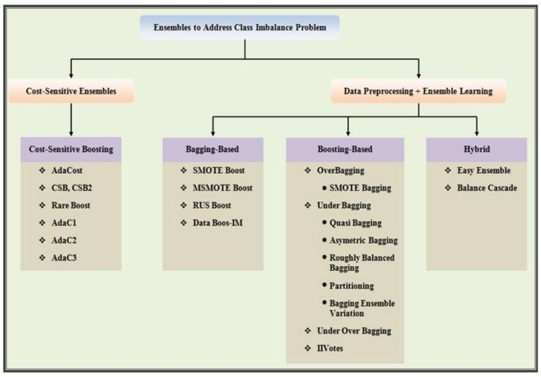

Figure 2: Different ELM-based approaches for imbalance class problems.

Hossam *et al.* [30] ejected a hybrid approach based on ensemble approaches. Feature selection, data balancing, along with classification were the ‘3’ stages in this model. In the prediction process, the influencing factors were identified initially. Then, a least powerful oversampling technique named SMOTE was utilized to handle the imbalanced data distribution. Regarding accuracy, Area under Curve (AUC), along with Geometric-Mean (G-Mean), better prediction outcomes were obtained. But, the outlier’s presence made the outputs unstable.

Yanjiao *et al.* [31] suggested a Parallel one-class ELM (P-ELM) centered Bayesian methodology for solving an imbalanced classification issue. Grounded on the class attribution of samples, the dataset was partitioned into k components here, which was further classified by subjecting it to the corresponding k Kernel-grounded one-class ELM classifiers. The suggested model's superiority was validated by the experiential analysis which was utilized in both binaries along with multiclass classification. However, the samples lost their intrinsic property owing to the sample area's overlapping.

2.3 Semi-Supervised ELM-based imbalanced classification problem.

Table 2 details the semi-supervised ELM-based techniques, their purpose, results, and limitations.

Table 2: Analysis of imbalance classification based on semi-supervised ELM

Authors	Method used	Purpose	Results	Limitation
Zhiqiong *et al.*[32]	Distributed and weighted ELM (DW-ELM) algorithm.	Primarily, each class's data center was engendered randomly. Grounded on the multivariate Gaussian distribution, the data of each class was engendered, where the formerly produced center point was pondered as the mean whereas the variance was the number of classes' reciprocal. Finally, the experimental data was made by combining all classes' generated data.	The approach’s efficacy was validated by relatively stable training time.	The HLs dimensionality was augmented by elevating the number of hidden nodes resulting in the algorithm's increased computation time.

Junjie *et al*.[33]	Self-Training Regularized Weighted Online Sequential ELM (ST-RWOS-ELM)	Initially, the positive and negative sample ratio along with the data chunk was unaltered. Also, by applying the weighting procedure, the errors in the samples were totally rebalanced.	Short calibration time along with low computational complexity and better accuracy of about 90%.	Imbalance classification was executed only for a highly labeled dataset.
Nan *et al*.[34]	UnSupervised ELM with Kernel (US-KELM)	The regularised cost function was diminished by utilizing a wavelet kernel function rather than output HL.	To attain better performance, it was capable of utilizing unlabeled data.	The training time was elevated by utilizing '2' loops in the kernel matrix computation.
Yongxiang *et al*.[35]	Laplacian-Hessian regularization SS-ELM (LHRSS-ELM)	Primarily, Laplacian and Hessian functions were incorporated with semi-supervised learning methods. The ELM's performance in semi-supervised tasks was elevated by a weighted assembled regularization approach via the sigmoid activation function.	The extrapolating power, accuracy, along with robustness in multiclass classification tasks were elevated by the combination of Laplacian and Hessian terms.	When the function was heavily oscillating, it was inaccurate.
Jin *et al*.[36]	Distributed Semi Supervised-ELM (DSS-ELM) with Zero Gradient Sum (ZGS) optimization strategy.	Each node in the communication network has an identical basic function and random parameters. The globally optimal coefficient vector was computed by ZGS strategy utilizing an iterative process	Guaranteed privacy-conserving algorithm with enhanced convergence	Disconnection might occur in communication nodes that cause data loss
Hongyu *et al*.[37]	Cost-Sensitive-Dissimilar-ELM(CS-D-	Grounded on the output's uncertainties acquired as of the	This scheme was insensitive to the choice of both the	The algorithm was made invalid and unavailable by

	ELM)	basic classifier, testing samples were partitioned into numerous groups initially. After that, sample groups with small uncertainties were pondered. The basic classifier that predicted the samples with their labels was summed to the original training set. Via the chosen training algorithm, the classifier was restrained on the enlarged training set finally.	classifier training algorithms along with the specific representation forms of uncertainty.	adding up low fuzziness to the training samples.

Carlos *et al.* [38] developed a Semi-Supervised Online Elastic ELM (SSOE-ELM) scheme to overcome the CI issues. During the training phase, each dataset was bifurcated into labeled and unlabelled instances. Grounded on the chunk utilized, training was executed in an online manner. Utilizing classification accuracy's mean and standard deviation as performance metrics, the experiment has recurred 30 times for each division. Especially in a situation like a relatively low amount of labeled data, SSOE-ELM was pondered as a feasible semi-supervised online learning algorithm. Nevertheless, the algorithm's accuracy was altered by the minute variation in the regularization parameter, as it was very sensitive.

Jie Yanga *et al.* [39] projected a Regularized Correntropy criterion-centered SSELM (RC-SSELM) technique. The correntropy was employed in the cost function's formulation. Non-Gaussian noises and outliers were dealt with by adopting the Maximum Correntropy Criterion (MCC) in the RC-SSELM's optimization strategy. A second-order statistical resemblance measure in the kernel space was the Correntropy. Thus, it provided better performance for a high percentage of outliers along with founding robust. However, poor performance was exhibited while readily propagating the misclassified outcomes in the labeled training data to the adjacent samples.

Parsa *et al.* [40] established online ensembles from multi-class imbalanced data for learning. Grounded on individual recall rates, the minority instances were over-sampled initially by the Improved SMOTE Online Ensembles (ISOE) while dynamically re-sampling whole classes. To re-balance the sample sizes of various classes, Improved Online Ensembles (IOE) were utilized by altering the parameter rate grounded on performance along with each class's number of instances. The unknown labels were processed by combining self-training into the online learning process. Thus, a better outcome was provided by the presented approach than the conventional schemes regarding balancing the predictive accuracies. Utilizing only static data sets was a major drawback here; the system's durability gets affected by the streams that do not contain concept drifts.

Zheng *et al.* [41] introduced a Semi-Supervised Broad Learning System (S2-BLS) for solving CI issues. Here, the mapped features were obtained by utilizing the ELM-grounded Auto Encoder (ELM-AE). Consequently, the discriminative projecting weights betwixt ground truths and the changed features encompassing the mapping features along with the enhancement nodes were computed. Subsequently, training the discriminant linear mode with altered samples along with ground truths occurs. Betwixt the labeled along with unlabelled samples, the information was explored efficiently. But, the training time was elevated by the useless information comprised in the retrieved mapped features along with enhancement nodes.

Adnan *et al.* [42] presented a Multi Kernel Semi-Supervised ELM (MKSSELM) method to overcome the unbalanced data conflicts. Utilizing the multi-kernel approach, the kernel parameters were fine-tuned initially. The kernel regulation was controlled by commanding a norm constraint on the kernel combination weights and optimizing the ELM structural parameters along with the kernel combination weights. The system's generalization performance was augmented by the Kernelized ELM. However, the local optima in the dataset affected the presented approach's performance.

Chuangquan Chena *et al.* [43] enumerated an Extreme Semi-Supervised Learning (ESSL) scheme to overcome imbalance classification. Initially, both binary, as well as multiclass classification issues, were handled by ELM in a unified model. Then, the HL was encoded by a tiny Approximate Empirical Kernel Map (AEKM) that diminished the computational cost along with memory usage. Subsequently, via the weighting function's elaborative design, the balance constraint or prior knowledge in the unlabeled data was eliminated. The imbalanced class problem was effectively along with efficiently solved by the ESSL without any fine-tuning parameters. However, the scalability was affected by the ESSL's dependence on regularization parameters.

2.4 Kernel- Fuzzy-based ELM imbalanced class problem.

Table 3 illustrated the diverse kinds of existing kernel-Fuzzy-based ELM approaches for imbalanced class problems.

Table 3: Evaluation of imbalance class problem based on kernel and fuzzy based ELM.

Researcher	Techniques	Procedure	Performance measure	Disadvantage
Weipeng *et al.*[44]	Fuzziness-centered Online Sequential ELM (FOS-ELM).	Utilizing the fuzzy classifier, meaningful data was chosen as of the given data initially. Here, the sequential learning utilized only the selected samples with high-output fuzziness.	Higher generalization performance with high testing accuracy.	The loss of vital information was caused by the utilization of predefined conditions.
Lu Li *et al.* [45]	AdaBoost Weighted Composite Kernel ELM (AdaBoost-WCKELM)	Grounded on the AdaBoost algorithm, the training sample's weights were adjusted here. The spatial information with original spectral bands was embedded by extracting the local mean and standard subsequently. Finally, the imbalance issue was solved by applying the composite kernel method in the ELM	Better accuracies were obtained even though the training samples were imbalanced or very limited.	Insufficient discriminative information for multiclass was provided by utilizing unlabelled data

		algorithm.		
Ping *et al.*[46]	(MKELM)-centered random forest binary classifier (MKELM-RF)	Multiple kernel functions comprised of different sub-feature spaces were utilized for space mapping. To tackle the CI issue, synthetic data points for the minority samples were generated by utilizing SMOTE. Finally, the features were chosen by executing bootstrap sampling to train MKELM-centered learners.	accuracy 98.1%, average positive predictive value 93.9%, and sensitivity 94.4%	Poor minority class sample detection.
Bhagat *et al.*[47]	Reduced Kernelized WELM (RKWELM)	To maintain equal instances as of each class, a subset of training data was utilized as the centroid of those kernel functions chosen by the ensemble method. The balance data was chosen by adapting the random under-sampling method.	A significantly lower running time when analogized with traditional methods.	A significant augmentation in the computational complexity with elevated dataset size was resulted by utilizing training samples for the kernel matrix.
Bhagat Singh Raghuwanshia *et al.* [48]	Generalized Class Specific Kernelized ELM(GCS-KELM)	The class proportion along with the regularization parameter's value was determined by the Class-specific regularization parameters. The input data was converted into feature space by utilizing the Gaussian	As it does not need the weight allotment to each training sample, the computational cost was diminished along with an accuracy of 92%.	The overfitting problem was raised in the approach.

		Kernel function.		
Bhagat *et al*.[49]	Reduced Kernelized-WELM(RK-WELM)	The CI issue was handled by utilizing an ensemble-centered solution initially. The base classifier used here was the RKWELM which generated balanced subsets on a training dataset. Finally, utilizing majority voting along with soft voting, those	Owing to the diminished number of kernels utilization, rapid classification along with 80% accuracy was obtained	Elevated data complexity resulted from choosing centroid kernels.

		base classifiers outcomes were combined.		

Shuya *et al.* [50] presented a Weighted Online Sequential ELM with Kernels (WOS-ELMK) for CI learning. To tackle the CI issue, the initialization phase, decremental learning phase, along with sequential learning phase was utilized in both binary class as well as multiclass datasets. Further, the non-optimal hidden node issue linked with random feature mapping was avoided by utilizing kernel mapping in WOS-ELMK. Thus, by getting not being adapted to the new data, the instability was avoided in the presented approach thereby elevating the presented approach's performance. However, the approach was made unsuitable for large datasets with the fixed memory scheme utilization.

Bhagat *et al.* [51] presented Balance Cascade-centered KELM (BCKELM) to tackle the CI problem. The training subsets were balanced by adapting a random under-sampling model. Further, the base learner was engendered by the ensemble method sequentially. A new balance training subset was engendered in every iteration. Therefore, better stability along with good generalization performance was provided by the presented approach. However, the computation time was elevated by the balanced training subsets dependence on the IR.

Hongyun *et al.* [52] developed the K-means SMOTE method along with the ANN-centered KELM algorithm (K-mean SMOTE+KELM) for solving CI problems. Utilizing the KELM approach, the data was balanced by integrating the K-means SMOTE technique with the ANN. Data weighting operation was executed by utilizing Deep Weighted ELM (DWELM) through an enhanced multiclass AdaBoost imbalanced learning algorithm (AdaBoost-ID) and an enhanced deep representation network centered ELM (EH-DrELM). The imbalance issue was sorted along with features' learning capacity was elevated by the K-mean SMOTE+KELM algorithm. However, the accuracy was affected by the value of the allocated weights.

Yang et al. [53] introduced Enhanced Kernel-centered Multilayer Fuzzy Weighted ELM (EML-KFWELM) to solve the CI problem. The minority class's classification performance was elevated by embedding a weighted strategy into ML-KELM by ML-KFWELM. Then, the classification errors caused by outliers and noise samples were eliminated by the fuzzy membership. Moreover, the ML-KFWELM's parameters were optimized by the EGWO strategy. The classification performance, accuracy along with stability was elevated by the optimization approach than other models. However, a slow response was made in the system owing to the existence of noise.

Zhennao *et al*. [54] addressed a hybrid ML model grounded on enhanced Grey Wolf Optimization (GWO) along with KELM. The GWO algorithm's global along with local search capabilities was elevated by a new hierarchical mechanism initially. The key parameters for KELM were tuned by utilizing IGWO. Therefore, a better convergence rate along with performance elevation was indicated by the IGWO-KELM's experiential outcomes. However, the computational burden was elevated by utilizing a discrete optimization strategy.

Pattaramon *et al.* [55] projected Radial Basis Function Network (RBFN) to tackle the class overlapping problem in imbalanced data. In instances where a common space was shared by more than one class, class overlaps occur generally. Additionally, shifting the decision boundary towards the negative class led to positive instances of misclassification near the class boundary. It cannot be utilized in a real-world application even though elevated classification accuracy was presented by this approach.

El Barakaz Fatima *et al.* [56] established three algorithms namely RONS (Reduce Overlapping with No-sampling), ROS (Reduce Overlapping with SMOTE), and ROA (Reduce Overlapping with ADASYN) to solve the CI problem. The classification of class-imbalanced data was optimized by diminishing the overlap degree along with looping the learning to separate datasets. However, the high dimensionality affected the presented system's performance.

Everlandio *et al*. [57] exaggerated evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning (EVINCI) to tackle the multi-class imbalance problem. A set of samples evolved as of an imbalanced dataset by utilizing a Multiobjective Evolutionary Algorithm (MOEA). In the overlapping regions, while choosing samples that generated precise models, the concentration of majority classes' fewer representative instances was opted. Regarding predictive accuracy, the approach's superiority was evaluated by the experiential outcomes. But the performance was affected by the larger IR.

2.5 Various other ELM-based approaches in imbalance classification.

Mingjing *et al.* [58] introduced a KELM parameter centered on a swarm intelligence algorithm called GWO. By exploring this for optimal parameter prediction, the KELM classifier's generalization capability was maximized by the swarm intelligence algorithms. Owing to the presented approach's superiority, it was utilized in the bankruptcy prediction. However, for datasets containing a large number of samples, it was not suitable.

Dong *et al.* [59] engendered an efficient bankruptcy prediction model centered on the KELM method. The optimal parameters were searched by a two-step grid search strategy that unites coarse search with fine search. The prediction was executed by the obtained optimal parameter finally. When analogized with the prevailing models, a better performance was exhibited by the presented model. However, a large variation in the classification accuracy was engendered by randomly assigned weights in distinct trials.

Yan Wei *et al.* [60] projected effective hybrid Gaussian Barebones (GB) improved Harris Hawk's Optimizer (HHO) centered KELM (GBHHO-KELM) method for predicting imbalance classification issues in students' intentions on self-employment. The global along with local search capabilities was balanced by introducing the GB mechanism into the HHO algorithm. Thus, better parameter combinations along with higher prediction sensitivity with more stable performance were attained by the presented scheme. However, the accuracy was affected by accomplishing the smaller fitness together with variance by the presented model.

Yongshan *et al.* [61] established a new Instance Cloned ELM (IC-ELM) in order to compute the class label of a testing instance. The testing instance's k nearest instances were selected by introducing the instance cloning technique. In the extended data set, each training instance's weight was computed. The underlying class label for the testing machine was predicted by formulating the learning model with an extended training dataset. Thus, the overfitting issue was more effectively handled by the presented model than the prevailing models. However, it cannot be utilized in practical applications like disease diagnosis, etc.

Xiaowei *et al.* [62] developed a Genetic Algorithm-centered ELM approach for the CI problem. The GA's population diversity was elevated by the Extinction and Immigration (EI) strategy. Here, the feature selection process utilized the Error Minimized-ELM (EM-ELM).

As per the ELM's generalization theory, an effective ranking method was formed for the ELM's selection process. Finally, the selected ELMs computed the final ensemble. Grounded on higher prediction accuracy along with better stability, the efficacy was demonstrated by the experiential outcomes. However, the system was made more time-consuming by the filter and embedded ensemble methods than the other schemes.

Ali Asghar *et al.* [63] introduced the Opposition Based Learning Grey Wolf Optimization (OBLGWO) algorithm for solving the CI problem. This algorithm included a greedy selection operator and LF mechanisms, random leaders, along with the strategy of random spiral motions. Regarding the faster convergence rates, the outcomes revealed the presented approach's better solution quality. The model selection issue was tackled effectively. However, the scalability was affected by the elevation in the average error and standard deviation that was augmented by the population dimension.

Xu Xiaolong *et al.* [64] presented a Density-based DSMOTE. Initially, the minority class's samples were partitioned into core samples, borderline samples, and noise samples by the optimized DBSCAN. To synthesize more effective samples, the minority class's noise samples were eliminated. Thus, the core samples along with borderline samples were oversampled by utilizing distinct strategies. Regarding precision, recall, and F-value, a better outcome was demonstrated by the experiential outcomes. However, the majority sample's information was not properly utilized and the presented model was susceptible to noise data.

Author name	**Approach**	**Steps involved**	**Advantage**	**Drawbacks**
Yang Lu *et al.* [65]	Individual Bayes Imbalance Impact Index (IBI^3)	The imbalance problem of a single minority class sample was measured by IBI^3. The degradation degree in the imbalanced data set was determined by BI^3.	Obtained higher correlated data with high F-scores.	The imbalance issue was not solved effectively since each type of classifier had distinct sensitivity.
Zhining *et al.* [66]	Self-paced Ensemble (SPE) approach.	The concern in classifying the classifier's samples was tackled by a classification hardness strategy. Here, better outcomes were attained by the resampling strategy which was guided by this hardness approach.	Robust performance even under highly overlapping classes.	The performance was affected by the computational cost's dependency on the dataset.
Mugdha *et al.* [67]	Convolutional Neural Network-based ELM	Initially, filters were used to learn the intuitive features like edges and basic shapes. More abstract features like texture were	By diminishing the overfitting problem, the translation variance was elevated by the CNN's utilization.	Some issues like elevated error rate and instability in classification were generated by the imbalance.

		acquired followed by features classification.		
Linbin *et al*.[68]	CI loss	The imbalance degree (ID) in the process of gradient descent was alleviated by a new loss function that could be employed among arbitrary imbalanced datasets.	Elevated recognition rate.	The efficiency was affected by the presented method's dependency on the number of samples.
Borowska *et al.* [69]	A Rough Granular computing (RGA)Approach	The granules with parameters were introduced initially. Utilizing the RGA approach, the parameters were tuned. RGA involved pre-processing followed by selective oversampling.	Small disjuncts, class overlapping, along with noises were carefully removed.	The complexity was elevated by the tendency to over-generalization and variance.
Adamu *et al.* [70]	Generative Adversarial Network(GAN)	A fine-grained generation was assured by utilizing multiple fake classes along with minority class instance's classification. Additionally, minority class instances aimed at rebalancing were also engendered.	Superiority regarding classification performance along with generated sample quality.	The performance was affected by the formation of poor minority class samples.
Xiaofen *et al.* [71]	Weighted ELM centered on Hybrid Artificial Bee Colony (WELM-HABC)	The ABC algorithm optimized the input bias initially. The HABC optimized the weights allocated to the training samples. The perturbed parameter vector's diversities of differential evolution	Superior classification performance with efficient AUC.	Failed to work in multi-class datasets.

		combine with the ABC's best solution information in the HABC effectively.		
Fatima *et al.* [72]	Multilayer Perceptron (MLP) and ELM.	The Credit card transaction's vector features were comprised in the input layer initially. The weighted inputs as of the input layer were received by the first HL that forward the earlier layer's data to the subsequent one. The classification outcome was contained in the output layer finally.	New fraudulent transactions were predicted rapidly with 95.46% accuracy.	The performance was affected by the false alarm's presence.

2.6 Performance analysis

The ELM algorithm's performance is validated here by means of a CI problem. Various models like SMOTE-CSELM [11], UBKELM [12], Ensemble-based WELM [17], and Demand Response Algorithm OSCSELM [13] were utilized by the prevailing schemes for the CI issue in crime scenes. Regarding Geometric Mean (G-Mean), Accuracy, and IR, the performance analysis are exhibited in Figures 3, 4, and 5.

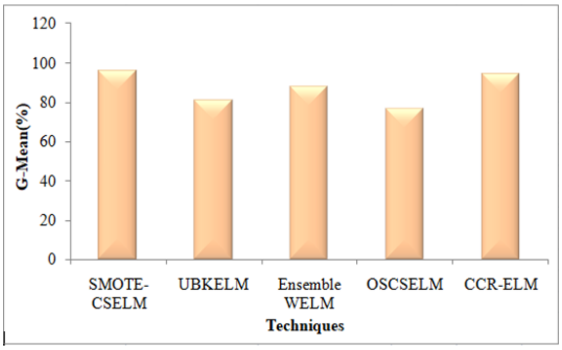

Figure 3: Performance evaluation based on G-Mean.

The model's G-mean is depicted in figure 3. 96.02% G-Mean was attained by the SMOTE-CSELM method whereas the UBKELM obtained only 81%. Similarly, 88.13%, 76.87%, and 94.50% G-mean were shown by the Ensemble-based WELM, OSCSELM, and the CCR-ELM method. Here, the SMOTE-CSELM approach is identified as more efficient than the other methodologies.

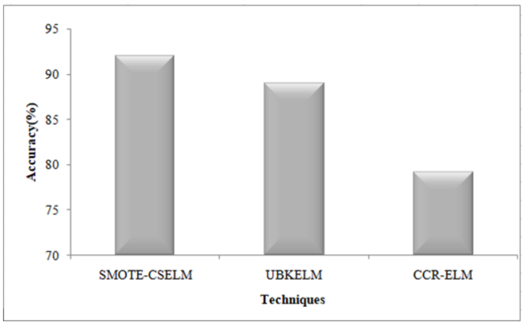

Figure 4: Performance assessment based on accuracy.

Grounded on accuracy rate, the performance of diverse models utilized in imbalance classification was revealed in figure 4. The number of appropriately predicted data points out of all data points is defined as Accuracy. Here, SMOTE-CSELM's accuracy is higher than the prevailing schemes. The SMOTE-CSELM's accuracy is 92%, whereas lower accuracy was attained by the CCR-ELM than the prevailing schemes; the UBKELM method's accuracy is 89%, which is much better than the CCR-ELM having 79.22%. It is concluded that better outcomes were attained by the SMOTE-CSELM than other approaches.

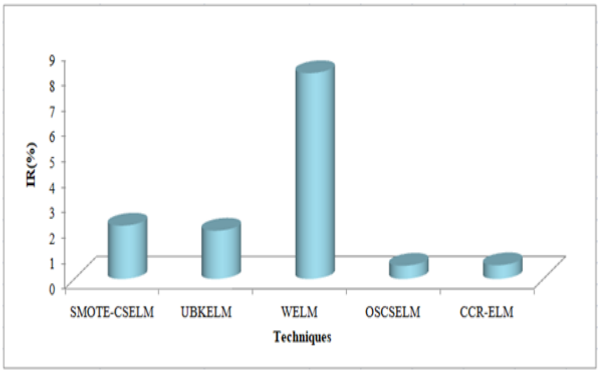

Figure 5: Performance evaluation based on Imbalance Ratio.

Regarding IR, other prevailing model performance evaluations were depicted in figure 5. The ratio of the number of examples belonging to the minority class to the majority class is named IR. The imbalance problem is higher if the IR value is smaller. It is evident that a lower IR value was exhibited by the OSCSELM method (0.52%) inflicting large imbalance issues in the dataset. Also, 2.1%, 8.11%, and 1.90% were the IR of SMOTE-CSELM, UBKELM, and ensemble-based WELM, respectively.

3 Conclusion

A significant issue in ML is classification with imbalanced class distributions. Regarding this topic, more related studies along with findings have been presented. A research review on several extreme ML techniques for classifying imbalances is presented here, along with a brief discussion on their drawbacks. Various prevailing models of extreme ML approaches for imbalance classification like Ensemble-based ELM for CI problems, Semi-supervised ELM algorithms for CI problems, and Kernel-based ELM for CI problems was explained by this literature work. Grounded on G-Mean, Accuracy, and IR performance indicators, the techniques' performance is examined. Finally, superior outcomes were exhibited by SMOTE-CSELM-based approaches than other methods that were concluded from this review. This review article recommends that future studies focus on making the system more reliable for imbalance classification via optimization and hybridized ML enhancements.

Declarations

Funding: No funds, grants were received by any of the authors.

Conflict of interest: There is no conflict of interest among the authors.

Data availability: All data generated or analysed during this study are included in the manuscript.

Code availability: Not applicable.

Author’s contributions: All Author is contributed to the design and methodology of this study, the assessment of the outcomes and the writing of the manuscript.

References

[1] Haseeb Ali, Mohd Najib Mohd Salleh, Rohmat Saedudin, Kashif Hussain and Muhammad Faheem Mushtaq. Imbalance class problems in data mining a review. Indonesian Journal of Electrical Engineering and Computer Science, 14(3):1560-1571, 2019.

[2] Paula Lauren, Guangzhi Qu, Feng Zhang and Amaury Lendasse. Discriminant document embeddings with an extreme learning machine for classifying clinical narratives. *Neurocomputing*, 277:129-138, 2017.

[3] Li Li, Ruizhi Sun, Saihua Cai, Kaiyi Zhao and Qianqian Zhang. A review of improved extreme learning machine methods for data stream classification. *Multimedia Tools and Applications*, 78(24):33375-33400, 2019.

[4] Jinhong Huang, Zhu Liang Yu and Zhenghui Gu. A clustering method based on extreme learning machine. *Neurocomputing*, 277:108-119, 2017.

[5] Jian Wang, Siyuan Lu, Shui-Hua Wang and Yu-Dong Zhang. A review on extreme learning machine. *Multimedia Tools and Applications*, 80(11-13):1-50, 2021.

[6] Peter Adeniyi Alaba, Segun Isaiah Popoola, Lanre Olatomiwa, Mathew Boladele Akanle, Olayinka Ohunakin, Emmanuel Adetiba, Opeoluwa David Alex, Aderemi A. A Atayero and Wan Mohd Ashri Wan Daud. Towards a more efficient and cost-sensitive extreme learning machine a state-of-the-art review of recent trend. *Neurocomputing*, 350:70-90, 2019.

[7] Huimin Pei, Kuaini Wang, Qiang Lin and Ping Zhong. Robust semi-supervised extreme learning machine. *Knowledge-Based Systems*, 159:203-220, 2018.

[8] Xiong Luo, Changwei Jiang, Weiping Wang, Yang Xu, Jenq-Haur Wang and Wenbing Zhao. User behavior prediction in social networks using weighted extreme learning machine with distribution optimization. *Future Generation Computer Systems*, 93:1023-1035, 2018.

[9] Kemal Akyol. Comparing of deep neural networks and extreme learning machines based on growing and pruning approach. *Expert Systems with Applications*, 140:1-14, 2019.

[10] Parashjyoti Borah and Deepak Gupta. Unconstrained convex minimization based implicit Lagrangian twin extreme learning machine for classification (ULTELMC). *Applied Intelligence*, 50(4):1327-1344, 2020.

[11] Bhagat Singh Raghuwanshi and Sanyam Shukla. SMOTE based class-specific extreme learning machine for imbalanced learning. *Knowledge-Based Systems*, 187:1-46, 2019.

[12] Bhagat Singh Raghuwanshi and Sanyam Shukla. Class imbalance learning using underbagging based kernelized extreme learning machine. *Neurocomputing*, 329:172-187, 2018.

[13] Sanyam Shukla and Bhagat Singh Raghuwanshi. Online sequential class-specific extreme learning machine for binary imbalanced learning. *Neural Networks*, 119:235-248, 2019.

[14] Hualong Yu, Xibei Yang, Shang Zheng and Changy. Active learning from imbalanced data a solution of online weighted extreme learning machine. *IEEE Transactions on Neural Networks and Learning Systems*, 30(4):1088-1103, 2018.

[15] Wendong Xiao, Jie Zhang, Yanjiao Li, Sen Zhang and Weidong Yang. Class-specific cost regulation extreme learning machine for imbalanced classification. *Neurocomputing*, 261:70-82, 2016.

[16] Bhagat singh Raghuwanshi and Sanyam Shukla. Class-specific extreme learning machine for handling binary class imbalance problem. *Neural Networks*, 105:206-217, 2018.

[17] Yong Zhang, Bo Liu, Jing Cai and Suhua Zhang. Ensemble weighted extreme learning machine for imbalanced data classification based on differential evolution. *Neural Computing and Applications*, 28(12):259-267, 2016.

[18] Hui Li, Xi Yang, Yang Li, Li-Ying Hao and Tian-Lun Zhang. Evolutionary extreme learning machine with sparse cost matrix for imbalanced learning. *ISA Transactions*, 100:198-209, 2019.

[19] Chengbo Lu, Haifeng Ke, Gaoyan Zhang, Ying Mei and Huihui Xu. An improved weighted extreme learning machine for imbalanced data classification. *Memetic Computing*, 11(1):27-34, 2017.

[20] Tianlei Wang, Jiuwen Cao, Xiaoping Lai and Badong Chen. Deep weighted extreme learning machine. *Cognitive Computation*, 10(6):890-907, 2018.

[21] Honghao Zhu, Guanjun Liu, Mengchu Zhou, Yu Xie, Abdullah Abusorrah and Qi Kang. Optimizing weighted extreme learning machines for imbalanced classification and application to credit card fraud detection. *Neurocomputing*, 407:50-62, 2020.

[22] Arnis Kirshners, Sergei Parshutin and Henrihs Gorskis. Entropy-based classifier enhancement to handle imbalanced class problem. *Procedia Computer Science*, 104:586-591, 2017.

[23] Xiaoyan Zhu, Chenzhen Ying, Jiayin Wang, Jiaxuan Li, Xin Lai and Guangtao Wang. Ensemble of ML-KNN for classification algorithm recommendation. *Knowledge-Based Systems*, 221:1-10, 2021.

[24] Wentao Mao, Jinwan Wang and Zhanao Xue. An ELM-based model with sparse-weighting strategy for sequential data imbalance problem. *International Journal of Machine Learning and Cybernetics*, 8(4):1333-1345, 2016.

[25] Junhai Zhai, Sufang Zhang and Chenxi Wang. The classification of imbalanced large data sets based on MapReduce and ensemble of ELM classifiers. International Journal of Machine Learning and Cybernetics, 8(6):1009-1017, 2015.

[26] Shifei Ding, Nan Zhang, Jian Zhang, Xinzheng Xu and Zhongzhi Shi. Unsupervised extreme learning machine with representational features. *International Journal of Machine Learning and Cybernetics*, 8(2):1-9, 2015.

[27] Fulong Ren, Peng Cao, Wei Li, Dazhe Zhao and Osmar Zaiane. Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm. *Computerized Medical Imaging and Graphics*, 55:54-67, 2016.

[28] Zhenyu Wu, Wenfang Lin and Yang Ji. An integrated ensemble learning model for imbalanced fault diagnostics and prognostics. *IEEE Access*, 6:8394-8402, 2018.

[29] Guillem Collell, Drazen Prelec and Kaustubh R Patil. A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data. *Neurocomputing*, 275:330-340, 2017.

[30] Hossam Faris, Ruba Abukhurma, Waref Almanaseer, Mohammed Saadeh, Antonio M Mora, Pedro A Castillo and Ibrahim Aljarah. Improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning a case from the Spanish market. *Progress in Artificial Intelligence*, 9(1):31-53, 2019.

[31] Yanjiao Li, Sen Zhang, Yixin Yin, Wendong Xiao and Jie Zhang. Parallel one-class extreme learning machine for imbalance learning based on Bayesian

approach. *Journal of Ambient Intelligence and Humanized Computing*, 2018, https://doi.org/10.1007/s12652-018-0994-x.

[32] Zhiqiong Wang, Junchang Xin, Hongxu Yang, Shuo Tian, Ge Yu, Chenren Xu and Yudong Yao. Distributed and weighted extreme learning machine for imbalanced big data learning. *Tsinghua Science and Technology*, 22(2):160-173, 2017.

[33] Junjie Wang, Zhenghui Gu, Zhuliang Yu and Yuanqing Li. An online semi-supervised P300 speller based on extreme learning machine. *Neurocomputing*, 269:148-151, 2016.

[34] Nan Zhang and Shifei Ding. Unsupervised and semi-supervised extreme learning machine with wavelet kernel for high dimensional data. *Memetic Computing*, 9(2):129-139, 2016.

[35] Yongxiang Lei, Lihui Cen, Xiaofang Chen and Yongfang Xie. A hybrid regularization semi-supervised extreme learning machine method and its application. *IEEE Access*, 7:30102-30111, 2019.

[36] Jin Xie, Sanyang Liu and Hao Dai. Manifold regularization based distributed semi-supervised learning algorithm using extreme learning machine over time-varying network. *Neurocomputing*, 355:24-34, 2019.

[37] Hongyu Zhu and Xizhao Wang. A cost-sensitive semi-supervised learning model based on uncertainty. *Neurocomputing*, 251:106-114, 2017.

[38] Carlos A. S da Silva and Renato A Krohling. Semi-supervised online elastic extreme learning machine for data classification. *International Joint Conference on Neural Networks (IJCNN)*, 08-13 July 2018, Rio de Janeiro, Brazil, 2018.

[39] Jie Yang, Jiuwen Cao, Tianlei Wang, Anke Xue and Badong Chen. Regularized correntropy criterion based semi-supervised ELM. *Neural Networks*, 122:117-129, 2019.

[40] Parsa Vafaie, Herna Viktor and Wojtek Michalowski. Multi-class imbalanced semi-supervised learning from streams through online ensembles. *International Conference on Data Mining Workshops (ICDMW)*, 17-20 November 2020, Sorrento, Italy, 2020.

[41] Zheng Liu, Shiluo Huang, Wei Jin and Ying Mu. Broad learning system for semi-supervised learning. *Neurocomputing*, 444:38-47, 2021.

[42] Adnan OM Abuassba, Zhang Dezheng and Zahid Mahmood. Semi-supervised multi-kernel extreme learning machine. *Procedia Computer Science*, 129:305-311, 2018.

[43] Chuangquan Chen, Yanfen Gan and Chi-Man Vong. Extreme semi-supervised learning for multiclass classification. *Neurocomputing*, 376:103-118, 2019.

[44] Weipeng Cao, Jinzhu Gao, Zhong Ming, Shubin Cai and Zhiguang Shan. Fuzziness based online sequential extreme learning machine for classification problems. *Soft Computing*, 22(13-14):3487-3494, 2018.

[45] Lu Li, Chengyi Wang, Wei Li and Jingbo Chen. Hyperspectral image classification by adaboost weighted composite kernel extreme learning machines. *Neurocomputing*, 275:1725-1733, 2017.

[46] Ping Yang, Dan Wang, Wen-Bing Zhao, Li-Hua Fu, Jin-Lian Du and Hang Su. Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification. *Biomedical Signal Processing and Control*, 63:1-13, 2021.

[47] Bhagat Singh Raghuwanshi and Sanyam Shukla. UnderBagging based reduced Kernelized weighted extreme learning machine for class imbalance learning. *Engineering Applications of Artificial Intelligence*, 74:252-270, 2018.

[48] Bhagat Singh Raghuwanshi and Sanyam Shukla. Generalized class-specific kernelized extreme learning machine for multiclass imbalanced learning. *Expert Systems with Applications*, 121:244-255, 2018.

[49] Bhagat Singh Raghuwanshi and Sanyam Shukla. Classifying imbalanced data using ensemble of reduced kernelized weighted extreme learning machine. International *Journal of Machine Learning and Cybernetics*, 10(8):3071-3097, 2019.

[50] Shuya Ding, Bilal Mirza, Zhiping Lin, Jiuwen Cao, Xiaoping Lai, Tam V Nguyen and Jose Sepulveda. Kernel based online learning for imbalance multiclass classification. Neurocomputing, 277):139-148, 2017.

[51] Bhagat Singh Raghuwanshi and Sanyam Shukla. Classifying imbalanced data using BalanceCascade-based kernelized extreme learning machine. *Pattern Analysis and Applications*, 23(3):1157-1182, 2019.

[52] Hongyun Qin, Houpan Zhou and Jiuwen Cao. Imbalanced learning algorithm based intelligent abnormal electricity consumption detection. *Neurocomputing*, 402:112-123, 2020.

[53] Yang Wang, An-Na Wang, Qing Ai and Hai-Jing Sun. Enhanced kernel-based multilayer fuzzy weighted extreme learning machines. *IEEE Access*, 8:166246-166260, 2020.

[54] Zhennao Cai, Jianhua Gu, Jie Luo, Qian Zhang, Huiling Chen, Zhifang Pan, Yuping Li and Chengye Li. Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf

optimization strategy. *Expert Systems With Applications*, 138:1-29, 2019.

[55] Pattaramon Vuttipittayamongkol, Eyad Elyan and Andrei Petrovski. On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems*, 212:1-17, 2021.

[56] El Barakaz Fatima, Boutkhoum Omar, El Moutaouakkil Abdelmajid, Furqan Rustam, Arif Mehmood and Gyu Sang Choi. Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection application to fraud detection. *IEEE Access*, 9:28101-28110, 2021.

[57] Everlandio R. Q Fernandes and Andre C. P. L. F de Carvalho. Evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning. *Information Sciences*, 494:141-154, 2019.

[58] Mingjing Wang, Huiling Chen, Huaizhong Li, Zhennao Cai, Xuehua Zhao, Changfei Tong, Jun Li and Xin Xu. Grey wolf optimization evolving kernel extreme learning machine application to bankruptcy prediction. Engineering *Applications of Artificial Intelligence*, 63:54-68, 2017.

[59] Dong Zhao, Chunyu Huang, Yan Wei, Fanhua Yu, Mingjing Wang and Huiling Chen. An effective computational model for bankruptcy prediction using kernel extreme learning machine approach. *Computational Economics*, 49(2):1-17, 2016.

[60] Yan Wei, Huijing Lv, Mengxiang Chen, Mingjing Wang, Ali Asghar Heidari, Huiling Chen and Chengye Li. Predicting entrepreneurial intention of students an extreme learning machine with gaussian barebone harris hawk's optimizer. *IEEE Access*, 2017, https://doi.org/10.1109/ACCESS.2020.2982796.

[61] Yongshan Zhang, Jia Wu, Chuan Zhou and Zhihua Cai. Instance cloned extreme learning machine. *Pattern Recognition*, 68:52-65, 2017.

[62] Xiaowei Xue, Min Yao and Zhaohui Wu. A novel ensemble-based wrapper method for feature selection using extreme learning machine and genetic algorithm. *Knowledge and Information Systems*, 57(2):389-412, 2017.

[63] Ali Asghar Heidari, Rahim Ali Abbaspour and Huiling Chen. Efficient boosted grey wolf optimizers for global search and kernel extreme learning machine training. *Applied Soft Computing*, 81:1-57, 2019.

[64] Xu Xiaolong, Chen Wen and Sun Yanfei. Over-sampling algorithm for imbalanced data classification. *Journal of Systems Engineering and Electronics*, 30(6):1182-1191, 2019.

[65] Yang Lu, Yiu-Ming Cheung and Yuan Yan Tang. Bayes imbalance impact index a measure of class imbalanced data set for classification problem. *IEEE Transactions on Neural Networks and Learning* Systems, 31(9):3525-3539, 2019.

[66] Zhining Liu, Wei Cao, Zhifeng Gao, Jiang Bian, Hechang Chen, Yi Chang and Tie-Yan Liu. Self-paced ensemble for highly imbalanced massive data classification. *IEEE 36th International Conference on Data Engineering (ICDE)*, 20-24 April 2020, Dallas, TX, USA, 2020.

[67] Mugdha Jain, William Andreopoulos and Mark Stamp. Convolutional neural networks and extreme learning machines for malware classification. *Journal of Computer Virology and Hacking Techniques*, 16(17):1-16, 2020.

[68] Linbin Zhang, Caiguang Zhang, Sinong Quan, Huaxin Xiao, Gangyao Kuang and Li Liu. A class imbalance loss for imbalanced object recognition. IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2778-2792, 2020.

[69] Borowska K and Stepaniuk J. A rough–granular approach to the imbalanced data classification problem. *Applied Soft Computing*, 83(4):1-40, 2019.

[70] Adamu Ali-Gombe and Eyad Elyan. MFC-GAN Class-imbalanced dataset classification using multiple fake class generative adversarial network. *Neurocomputing*, 361:212-221, 2019.

[71] Xiaofen Tang and Li Chen. Artificial bee colony optimization-based weighted extreme learning machine for imbalanced data learning. *Cluster Computing*, 22(3):6937-6952, 2018.

[72] Fatima Zohra El hlouli, Jamal Riffi, Mohamed Adnane Mahraz, Ali El Yahyaouy and Hamid Tairi. Credit card fraud detection based on multilayer perceptron and extreme learning machine architectures. *International Conference on Intelligent Systems and Computer Vision (ISCV)*, 09-11 June 2020, Fez, Morocco, 2020.