Extreme Learning Machine Based Approaches For Imbalanced Data Problem

Ms. Shivani sharma[1], Dr. Nirupma tiwari[2]
[1, 2] Dept of computer science and engineering
[1, 2] SRCEM, RGPV university, Bhopal

Abstract- *There have been significant developments in recent years and changes in the classification of data have been made. The class imbalance problem in data mining is now the biggest. The problem of imbalance arises where one of the two classes has more samples than another. Most of the algorithms rely mainly on classifying the primary sample while ignoring the minority sample. The minority samples are uncommon but still significant. ELM consists of randomly assigned learning parameters of hidden nodes, comprising input weights & biases, and not changing them whilst analytically evaluating output weights by quick, generally appropriate inverse operation. This work initially presents a brief review of Imbalanced data and imbalanced data classification, also describing the data classification as well as classification methods such as- ID3, C4.5, KNN, naïve bayes, SVM, ANN algorithm, etc. Next, the paper summarized an overview of machine learning and machine learning techniques, also we deeply describe extreme learning machine (ELM) & different types of variants. In the last, brief description of SMOTE. A sensor network can be made scalable by forming clusters.*

Keywords- Data Mining, Imbalanced Data, Imbalanced Data Classification, Data Classification, Machine Learning, ELM, SMOTE.

I. INTRODUCTION

Data mining (DM) is a method of data extraction from large datagrams by using algorithms and technology derived from statistics, machine learning, and database management systems. Traditional methods of data analysis also require manual work and slow, costly, highly subjective analysis of data. Data mining, which is widely referred to as knowledge discovery in large-scale data, allows companies and organizations to take measured decisions through the assembly, compilation, analysis, and access to corporate data. It uses several methods, such as queries and monitoring tools, analytical processing tools, and DSS (Decision Support System). Anomaly detection is an important problem that has been researched within diverse research areas and application domains.

Classification is an instructional process tracked. The two-step method is data classification. In the first step, data multiples from training data with a set of attributes are evaluated to form a model. The value of class label attribute is defined for each tuple in the training data. The classification algorithm is used to build the model on data training data. Test data is utilized to validate the accuracy of the model in second step of classification. The algorithm will be used to identify unknown tuples if the accuracy of the model is appropriate. To derive rules & patterns from data that could be used for prediction, classification techniques are developed as an essential component for machine learning algorithms. Techniques of classification are used to classify data records in one of a set of predefined classes. They are developing a training dataset model consisting of examples with recognized class labels [1].

Imbalanced datasets are typically used in real-life applications. The data science community draws a lot of interest in mining such as databases and research efforts. The key explanation for this is that the use of standard classification algorithms for imbalanced data mining produces unreasonable outcomes, particularly if the datasets are highly imbalanced with many features. A standard learning algo will most likely misclassify a large portion of minority classes if the dataset were skewed towards minority class [2].

Machine Learning (ML) paradigm can be used to enhance future performance by learning from experience (in this case preceding data). Automatic learning methods are the only focus of this area. Learning is a modification or enhancement of algorithms based on past "experiences" without external human assistance. ML is an AI (Artificial Intelligence) branch that advances the notion that machines can learn by themselves how to solve a particular problem by having access to the right information. By using sophisticated mathematical and statistical tools, ML enables machines to carry out intellectual tasks independently, which humans have traditionally solved. This concept of automating complex tasks has produced a considerable interest in the area of networking in anticipation of uploading many operations in the design & application of communication networks to machines. In various network areas, some applications have already fulfilled these prospects in an area like traffic classification, cognitive radios, intrusion detection [3].

The extreme learning machine (ELM) has recently begun to gain traction in the field of transmission learning. ELM has played a part in the popularity of learning speed and fast implementation of the ELM algorithm. The simplicity of using sequential online learning methods at ELM was also a significant consideration. The purpose of this work is to provide a summary of the strategies used to integrate ELM & transfer learning in published works. It also focuses on the benefits & limitations of every algorithm whilst also recommending future research aspects of ELM transfer learning [4].

SMOTE (Synthetic Minority Oversampling Technique): Throughout this process, the minority classes are overrepresented by generating "synthetic" instances instead of over-sampling with substitution. It is a statistical method for balancing an increased amount of cases in the dataset. The module starts by making new files, delivered as input data, from available minority incidents. Such an application of SMOTE does not affect the number of majority incidents [5]. Several methods are available for the over-sampling dataset used in the standard grading problem (a classification algo is used for the classification of an image, provided a labeled images training set). The popular approach is called SMOTE: an over-sampling technique for the synthetic minority. For eg, certain training data that have samples and f functionalities in the data feature region would be included in this technique [5].

II. IMBALANCED DATA

Imbalanced data provide an unequal representation of data samples between classes which creates a challenge for learning algorithms as learning minority class principles become challenging. This issue is solved by synthetic oversampling approaches by developing synthetic minority samples to balance the data sets. Most of these approaches could therefore generate incorrect samples from synthetic minorities within majority areas [8]. Sampling takes the form of over- or under-sampling as the most common method for classifying unbalanced datasets. Both strategies adjust the distribution of data by removing class majority cases or increasing the class of the minority. If so, any data that might be useful may be deleted or over-sampled, the probability of over-sampling increases. SMOTE is a widely known method of producing new samples belonging to the minority class. A method is common & numerous aspects have been studied: the choice of samples, different methods of obtaining new data, the use of all samples, etc [6].

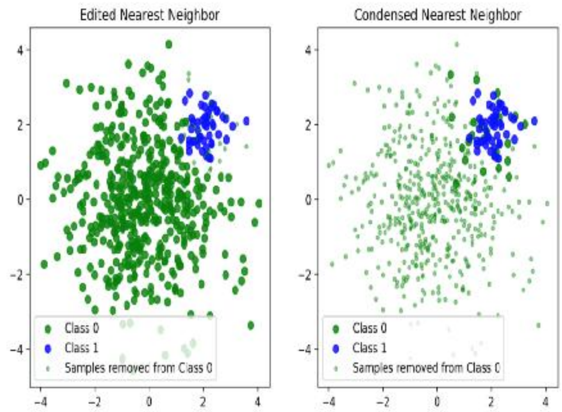

Figure 1. Imbalance Data

III. IMBALANCED DATA CLASSIFICATION

Imbalanced data classification is a concern with a comparatively wide margin in which proportional class sizes of datasets vary. In this class, at least a few numbers are seen in one class (named minority class) & the rest falls into other classes (named majority class). In essence, the performance of classifier in imbalanced data set is skewed against several classes (majority class). Performance bias means that in large and small groups solutions behave differently. Solutions in the majority class are more specific. However, the results are poorly reliable from a minority class point of view. In practical applications like bleeding detection in medical diagnoses, fraud detection, & so on, an issue of imbalanced data delivery is well represented. Safe level smote, neighborhood cleaning rule, neural network & cost-sensitive algorithm are the most common approach for imbalanced data. Figure 2 provides an example of imbalanced data [7].

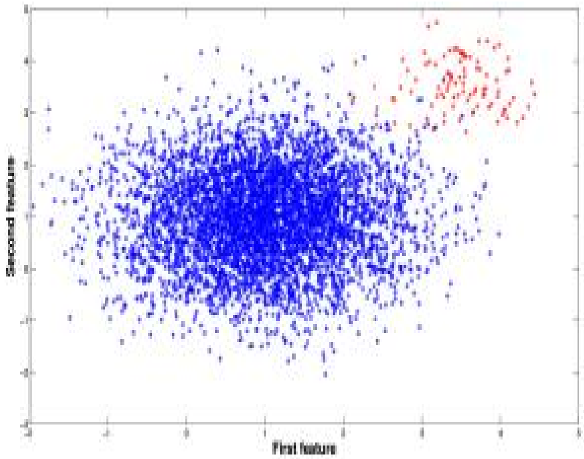

Figure 2. Example of imbalanced data problem

The data classification proposes 3 wide approaches to tackle imbalance issues: algorithmic-based approaches, hybrid approaches, and pre-processing approaches.

A. Approaches for Imbalanced Data Classification

Diverse methods to address the imbalanced data classification have been proposed over the last decade. This section offers a comprehensive analysis of state-of-the-art machine-learning approaches to solve this issue. This includes a review of pre-processing techniques, cost-sensitive methods of learning, hybrid methods, and algorithmic approaches. Figure 3 demonstrates the classification of different approaches to handling imbalance data.

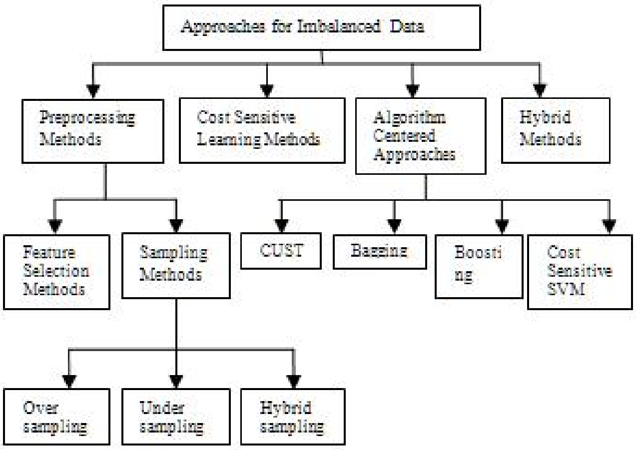

Figure 3. Classification of approaches for imbalanced data

- **Pre-processing Approaches**

These approaches those carried out on training data. Pre-processing methods are used to provide more data on training. Methods used in pre-processing are also known as data-centered approaches. These methods work by working specifically on data space & attempting to decrease imbalance ratio b/w classes.

Over-Sampling applies artificial samples to data space; named SMOTE, the most common technique. The following sections define the different methods of sampling.

- Sampling Methods
- Feature Selection and Extraction

- **Cost-Sensitive Learning Approaches**

These approaches are cost-specific methods that define cost-related examples. This typically requires exploration costs for misclassification. Considering that the misclassification costs do not demonstrate data and problems are caused in setting costs, cost-sensitive methods of learning become less popular than sampling techniques. In contrast with these, sampling techniques are easily implemented & most common. However, this learning approach is a more mathematically effective method.

- **Algorithm Centered Approaches**

The development of new algorithms or enhancement of the existing one is named algorithmic-based techniques to solve the imbalanced data classification problem. Researchers and literature for algorithmic-based techniques have suggested different methods.

- **Hybrid Methods**

Incorporating approaches of pre-processing & algorithm-oriented to handle the problem of imbalanced classification essentially leads to hybrid methods [8].

IV. DATA CLASSIFICATION

The classification of data is one of the most frequent strategies for DM whose role is to define the target class of which those objects are part of an unknown class. There are several conventional data mining classification algorithms. However, the advantages and drawbacks of each classification algorithm. The result of an algorithm of classification is generally related to data features. Among the various classification algorithms, a classification algorithm of the SVM (support vector machine) suggested by Vapnik et al. is widely used and depends upon the SLT principle for structural risk minimization. In solving problems of classification, such as small sample numbers or non-linearity, SVM has many special advantages. Traditional SVM classification algo brings 2 ideas on board: training size examples in a given collection of data are balanced and error cost in classification approximately equal. Nevertheless, several problems of imbalanced data classification arise in the real world. Moreover, the accuracy of the class of minorities is much more significant. As mentioned above, SVM's classification model is based on structural risk minimization. The examples of the majority class then have a major impact on classifiers for imbalanced results, therefore providing positive classification weight for the main class & seriously affecting the distribution of classification hyperplane. It is most significant, therefore, that classification techniques at any algo or data level can be enhanced to overcome the classification of imbalanced data, at present trending problem in the research field of DM. [9].

A. Data Classification – Basic Inductive Methods

5 methods of classification introduced in this work are discussed in this section. The analysis gives us a basis for

creating hypotheses about potential link b/w data characteristics & method performances. While several innovations to AI methods to solve particular problems have been made, only simple models are used in this analysis, to preserve genuine features of original algos. For a specific problem situation, AI & statistical approaches may be fine-tuned. Further calibrated model; however, harder it is for new problems to become generalized [9].

a. ID3 Algorithm

The original set as a root hub begins the ID3 calculation. It emphasizes the entropy (or IG (A) data collection) of the attribute by each cycle of the algorithm by each unused attribute in the set. The attribute with smallest entropy (or largest data gain) value is then chosen. Set is then separated into S by selected attribute (for example, marks < 50, marks < 100, marks > = 100) to generate information subsets. The algorithm restores each object of the subset and takes into consideration only items not previously selected.

b. C4.5 Algorithm

It is a decision tree algo used to expand the estimation of a previous ID3 calculation. It improves ID3 algo by handling properties of continuous & discrete, missing values & trees after creation, which are also regarded as a statistical classification, decision trees generated by C4.5 can be used to group. C4.5 generates decision trees with the same Id3 algorithms from a collection of training data. As a supervised learning algo, a series of training instances are required that can be considered as a pair: an object i/p & desired o/p value (class). Algorithm analyses and constructs a classifier capable of grouping training and test cases correctly.

c. K-Nearest Neighbor (KNN) Algorithm

Nearest neighbor (NN) rule differentiates classification of indefinite data point based upon its NN whose class is already recognized. M. Cover & P. E. Hart purpose KNN in which NN is measured based on k evaluation that shows how many NNs are to be studied to describe a class of sample data point. It makes use of above one NN to define a class in which provided data point belongs to & accordingly it is known as KNN.

d. Naïve Bayes (NB) Algorithm

NB Classifier approach depends upon the Bayesian theorem also is especially utilized when the dimensionality of inputs is high. Bayesian Classifier is up to estimating the most possible output based on the input. New raw data can also be applied at runtime and a better probabilistic classifier is available. The existence (or absence) of a specific feature (attribute) of a class is not associated with the presence or absence of any other feature in an NB classifier when a class variable is given.

e. Support Vector Machine (SVM)

SVM is depending upon the concept of statistical learning as well as structured risk reduction which attempts to identify the location of decision limits, also recognized as hyperplanes that generate the optimum class distinction. The utility of the classification based on SVM is not explicitly dependent on the classified entity dimension. Although the most effective and precise classification algorithm is SVM, as there are many issues. Analysis of data in SVM is dependent upon convex quadratic computing & seems to be costly computationally since large matrix functions and also time-consuming numerical computations are needed to resolve quadratic programming methods.

f. Artificial neural networks (ANN Algorithm

ANNs are types of computer architecture stimulated by BNN biological neural networks (Brain nervous systems) also are utilized to approximate functions that can be based on large no. of inputs also are usually unidentified. ANNs are proposed as interconnecting "neuron" structures that can compute input values and, because of their adaptive existence, can learn both from machine learning and pattern recognition. ANN works by forming relationships in the biological brain between various processing elements, every corresponding to a single neuron. These neurons may be created or replicated using a digital computer system. Every neuron receives multiple i/p signals and then generates a single o/p signal dependent on internal weighting, which is sent to another neuron as an input. Neurons are intertwined intensely & arranged into multiple layers. I/p layer is supplied with i/p and the o/p layer is generated. One or more hidden layers between the two are usually sandwiched.

V. MACHINE LEARNING (ML)

ML is a method utilized to show computers how to work with data more easily and to produce a better performance. We can't grasp the model or derive details from the data in certain situations after presenting the dataset. We use machine learning techniques in such cases to predict the results. A large number of datasets can be obtained from numerous sources, and machine learning is required. ML is used in several fields, from medical to military, to derive useful knowledge from accessible data sets. Machine learning is specifically meant to learn from current records. A wide variety of algorithms is programmed to learn from computers. Often mathematicians and programmers use a variety of methods to solve this problem [10].

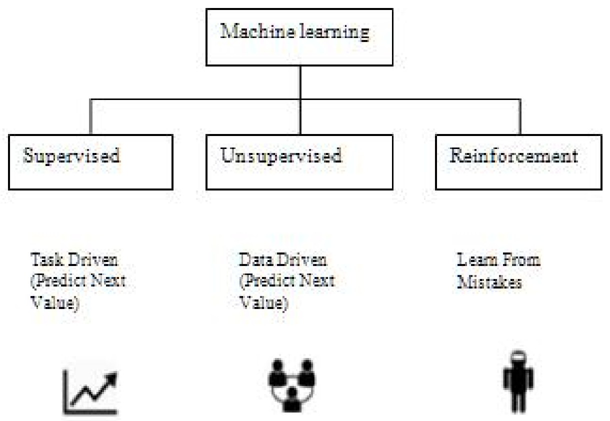

Figure 4. Machine Learning

B. Machine Learning Techniques

ML has been one of the key elements of Information Technology (IT). There is a motive to expect that smart data processing will become much more popular as a critical ingredient for technical purposes, as data becomes increasingly available. Most of the machine learning research is about solving these challenges and finding successful solutions to them. A problem may be modeled according to its interactions with experience or world or i/p data in various ways by an algorithm. In this entire, we must first adopt an algorithm which can adopt a learning style. The algorithm may have very few big learning models. The way ML algorithms are organized is useful as it allows one to understand the roles of i/p data and model of planning procedure & to choose what is ideally matched to the target results for the challenge. An algorithm for machine learning is called the recognition of an entity. Let us explore the various learning types and their different elements in machine learning algorithms [3]:

1) **Supervised learning:** It includes several pre-labeled and target data input variables (training data). It produces a mapping function for mapping inputs to the necessary outputs using the input variables. The sufficient precision of the system about the teaching data is obtained by the parameter adjustment process. Learning in many applications is used, like spam detection, object recognition, as well as speech recognition. The purpose is to estimate the value of one or more o/p variables given i/p vector value x. O/p variable may be either continuous (regression) or discrete (classification problem). A training data collection contains N samples of i/p and required o/p values [11].
2) **Unsupervised learning:** We only have training data in this algorithm rather than results. The input data is not labeled beforehand. In classifiers, the patterns or clusters in input data sets are known. In the case of unsupervised learning, only a set of input vectors x is used. Although unsupervised learning may manage numerous activities, the study of clustering or clusters is the most popular.
3) **Reinforcement learning:** A machine is trained to create such decisions using this algorithm. Based on any single point of data, these algorithms select an action and learn then how good the decision was. It is an environment where a machine is exposed [12]. In general, Reinforcement Learning (RL) is used as a method for approaching applications like robotics, financing (investment decisions), and inventory management. The goal is to learn a strategy that is to map environmental states into actions when interacting directly with the environment. RL model helps agents to research available behaviors and optimize their actions with only reliable feedback, called the incentive. The goal of the agent is to optimize its performance in the long term. The agent thus not only considers the immediate reward into consideration but measures the impact on the future of its actions. The two most critical characteristics of RL are delayed reward and trial/error [3].

VI. EXTREME LEARNING MACHINE (ELM)

For training of SLFNs (single hidden layer feedforward neural networks), ELM has been proposed. With ELM, random initiation, and then fixation of hidden nodes without iterative tuning. And neurons do not necessarily have to be hidden nodes in ELM. The connection (or weights) b/w hidden layer and o/p layer are only free parameters to be known. ELM is thus constructed as a linear model that reduces it to a linear system resolution. ELM is extremely effective and aims to achieve an optimal overall relative to traditional FNN learning approaches. Theoretical analysis has shown that ELM retains the universal approximation capacity of SLFNs even with arbitrarily produced hidden nodes. ELM can obtain virtually ideal generalization bound of typical FNN with regularly used activation features in which all its parameters are known. ELM's advantages over conventional FNN algorithms were demonstrated with a broad variety of problems in various fields in terms of efficiency and generalization performance. It is important to note that ELM is much more effective commonly than SVM, least-square SVM also other state-of-the-art algorithms. Empirical experiments have shown that ELM is comparable, or enhanced than SVMs & their variants in terms of generalization [13].

- **Variants of ELM**

This portion introduces and describes a few common ELM variants.

a) Incremental ELM (I-ELM)

I-ELM to create an incremental input network. In the secret layer, I-ELM inserted random nodes. This was added one at a time. This froze the performance weights of currently hidden nodes (HNs) when new HNs were inserted. In addition to the continuous (and differentiable) enabling functions of the SLFN, I-ELM is effective for SLFNs that have partly continuous (such as threshold) activation functions. Huang et al. introduced enhanced I-ELM (EI-ELM) & convex I-ELM (CI-ELM) given this I-ELM sense.

b) Pruning ELM

Originally, P-ELM starts with a huge no. of hidden neurons were randomly selected weight dimensions are assign to achieve its input vector response. Hidden layers are adjusted and preserved based on the value of each hidden node in evaluating the classifier's predictive performance. P-ELM uses statistical measurement to determine the degree of significance among the class labels and the hidden node. Then this eliminates the hidden reduced-relevant nodes to explore the compressed network without losing its generalization power. P-ELM starts with many HNs before less relevant or irrelevant HNs are eliminated by considering their importance in learning for class training samples.

c) Error-minimized ELM

This system is capable of increasing group by group hidden nodes and of knowing the amount of HNs in generalized SLFNs automatically. During network development, performance weights are changed slowly and thus the device complexity is significantly reduced. For secret nodes of sigmoid form, the results of the simulation indicate that this technique will dramatically reduce the computation difficulty of ELM & help build an efficient application of ELM.

d) Online sequential ELM

Both preparation details will be accessible for training purposes by utilizing traditional ELM. Training data can however be accessed one by one or one element per line in actual applications. Proposed an algo called the online ELM(OS-ELM) for sequential learning. This algo can be used in a unified environment with both RBF and additive nodes.

e) Ordinal ELM

ELM algo for common problems in regression, three ordinal algo based on ELM has been implemented and one framework is based on encoding.

f) Fully complex ELM

ELM algo was generalized in this application from real domain to abstract domain. Like ELM, C-ELM's secret layer preferences and input weights were randomly chosen based on the ongoing likelihood of spread. The performance weights were then measured analytically, rather than iterated. [14].

g) TELM (Twin Extreme Learning Machine)

The twin SVM is applied to extend the SVM, which aims to derive two non-parallel distinguishing hyperplanes. Each hyperplane enters one of its two classes at the shortest path which allows the distance as wide as possible for the other category. Besides solving two lowered-sized QPPs, it has quicker learners. In ELM feature space for the classification process, the TELM learns 2 nonparallel segregating hyperplanes. For each hyperplane, TELM is being utilized to reduce distance to one of two classes as well as maintain it far away from another class. TELM aims to reduce both error rate as well as the number of distance squares through one hyperplane to one of the two groups.

h) CCR-ELM (Class Specific Cost Regulation ELM)

Sometimes binary classification focuses on the classification techniques for extremely unbalanced outcomes. Multi-class classification eliminates the problem of an unbalanced set of data. Thus, it is hard to distinguish the minority and majority groups. To prevent such difficulties of CCR-ELM (class-specific cost control ELM) can be employed for classification difficulties with imbalanced class distributions. It establishes class-specific enforcement expenses for misclassification in the quality index context of each class. This minimizes the impact of no. of class samples as well as effects of degree of data dispersion.

i) ML-ELM (Multilayer Extreme Learning Machine)

It is a classification method utilized for classification of motor imagery EEG (Electroencephalogram). This is among the most significant BCI technologies utilized. Throughout this structure, the mixture of LDA and PCA is selected as the feature extraction techniques as well as the ML-ELM techniques are applied for the classifier. ML-ELM does better than ELM [15].

VII. SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)

SMOTE is an over-sampling technique that generates synthetic minority class samples. It is utilized to achieve a synthesized or almost class-balanced training set that can then be used for the training of the classifier. SMOTE is an over-sampling approach, which produces synthetic samples for the minority class. It can be better than just over-sampling and is commonly used. The over-sampling process requires the random copying of minority samples to maximize no. of minority class samples so that examples of minority and majority class sizes can be balanced. SMOTE is a very common tool for generating new data. It is dependent on sampling minority class data by modestly making segment data points that connect one of its nearest K-nearest data points to a randomly selected data point. It is a very basic method and has become widely common in use. SMOTE's only problem was that it is not founded on a sound theory in mathematics [16].

While SMOTE is not intended to imitate underlying distribution, distribution is important for the establishment of classification limits. We also explain SMOTE’s impact on classification results, as classification efficacy is the primary goal when utilizing SMOTE, in addition to our proposed distributional analysis. In reality, our priorities are the same [17]:

- Improve the mathematical model of SMOTE & calculate to what degree it emulates = underlying distribution (check its moments). = presented theory is universal, & is valid for any distribution.
- Put on terms of identity statistical analysis to 2 distributions: Gaussian or Multivariate Gaussian distribution to achieve simpler, closed-form for medium or covariant sequence distribution over-sampled.
- Include a thorough laboratory analysis of SMOTE, analyzing factors influencing its accuracy (imitating distribution). For instance, we find both statistically &empirically, the number of initial minority trends decreases, as scale grows, and as no. of neighbors applied to analyze SMOTE grows, the accuracy deteriorates.
- Analyze utility of SMOTE for other classifiers, both logically & empirically, by examining the impact of specific variables on their efficacy;
- Give a detailed analytical study of SMOTE including three common SMOTE delays (SMOTE1Borderline, SMOTE2Borderline, or Adasyn) to analyze the distribution or classification efficiency of this over-sampling method [17].

VIII. LITERATURE SURVEY

Wong, S. Y. [2020] examine ELM's ability to generate probabilistic performance from the existing ELM structure itself while maintaining ELM's merits without any need for two-stage post-processing operations to transform the performance to likelihood as well as remove iterative training to determine output values. Two unified probability-based ELM architecture methodologies are provided, such that, PO-ELM (Probabilistic Output Extreme Learning Machine) and CPP-POELM (Extreme Learning Machine based on Constrained Posterior Probabilistic Outputs). To demonstrate its effectiveness and validity in managing pattern optimization problem and also decision-making processes, the suggested algorithms are trained empirically on multiple benchmark functions and realistic power system applications [18].

R. Rastogi and A. Prasad [2019] A hybrid approach is proposed, accompanied by an Extreme Learning Machine, for classifying binary imbalanced data using SMOTE. Our way of predicting the desired class with a swift learning rate is efficient. Our model was tested using five standard imbalanced data sets and G-mean, F-measure & ROC score was higher for all of the datasets [19].

Z. Yuan and P. Zhao [2019] Data mining provides a widespread imbalance of data and ensemble algorithms are a more efficient classification to identify imbalance data. However, for imbalanced data, the ensemble learning algorithm itself is not optimized. Therefore, it is proposed an imbalanced method of data processing based on the data level of SE-gcForest ensemble learning. This approach incorporates the concept of processing data level after the gcForest algorithm’s multi-grained scanning window process. The Simple Ensemble concept and the SMOTE algorithm work with imbalanced data. Experiments demonstrate that when the data imbalanced ratio has been lower and higher, this approach is more efficient. [20].

D. Durga Prasad, **et al. [2019]** To discover information from the imbalance dataset, suggest an appropriate algo called WIMUS (With In Majority Under Sampling). WIMUS algo only utilizes intrinsic features of datasets to exclusively under-sample instances from majority subset. In 12 imbalanced datasets from UCI Repository, WIMUS algo is comparable with the SMOTE approach. The findings show that the method suggested is effective compared with precision, AUC, F-measure & recall approach [21].

Lee, Z.-J et al. [2019] An intelligent, distributed algorithm for imbalanced datasets is presented in this paper. In proposed intelligent distributed algorithm, Apache Spark is introduced as a dynamic network and its in-memory data processing engine cluster computing system that can evaluate a large amount of information. Throughout the distributed framework, Apache Spark is proposed to access unbalanced data first with SMOTE. After that, to distinguished imbalanced data, the SVM is utilized. To check the efficacy of the proposed algorithm, a zoo set of data from the UCI ML repository will be utilized. The performance of proposed intelligent distributed algo will achieve better performance than conventional classifiers as contrasted [22].

M. Lv, et al. [2019] The credit card data used by the UCI database is imbalanced consumption data. For the analysis of imbalanced data, the sampling method SMOTE, AdaBoost algorithm, and a cost-sensitive algorithm is used. Empirical studies propose that the SMOTE-AdaBoost technique is more effective than the conventional AdaBoost technique, and the cost-sensitive algorithm that raises the weight for minority class samples than the traditional AdaBoost method is also higher. This paper finally outlines the problems. [23].

L. Xiang, et al. [2018] Fix this issue with the proposed fast unsupervised heterogeneous data learning algo, called Two-stage Unsupervised Multiple Kernel Extreme Learning Machine (TUMK-ELM). Alternatively, TUMK-ELM extracts data by various sources & learns heterogeneous data representation via closed-form solutions that enable it to be quite fast. As theoretical evidence has demonstrated, at each stage TUMK-ELM possesses low computational complexity, and iteration of these 2 stages can be converged infinite steps. As proven on 13 real-life data sets, TUMK-ELM increases dramatically in performance relative to 3 state-of-the-art heterogeneous approaches for data learning (up to 140 000 times) with comparable efficiency. [24].

Y. Ge, et al. [2017] proposes an improved SMOTE algo (MBK-SMOTE) which depends upon the standard SMOTE algorithm. MBK-SMOTE makes samples distribution balanced by introducing the concept of sample's density & dividing concentration area. Then we use the Random Forest algorithm to predict the icing event of wind turbine blades. Comparing the experiments among MBK-SMOTE+RF, standard SMOTE+RF, and improved SMOTE+RF, we finally found that MBK-SMOTE+RF is more effective, and the prediction effect of MBK-SMOTE+RF is superior to other methods [25].

IX. CONCLUSION

We describe different approaches and innovations used in recent studies in this review paper. To investigate organized data and rules, data mining is a discovery and evaluation of vast collections of data. Machine learning is a computer science field where machines can learn from data without being programmed directly. The most prevalent issue is the data imbalance. The data imbalances are not consistent with current classification algorithms, so we have to plan and balance data. Sampling & cost-sensitive learning are techniques for allowing data balance. A sampling at the data level is the most frequent way of handling imbalanced data. Over-sampling is better than under-sampling for local classifiers, while several under-sampling approaches outperform over-sampling by using global learning classifiers. In this paper, a survey on the conceptual map of the most common imbalanced data and the imbalanced data classification approaches. Here we have looked at data classification and some data classification techniques. Also, a description of ELM and different variants of ELM has been stated. Although, conclude the machine learning and machine learning techniques, also described SMOTE in this paper.

REFERENCES

[1] Ashish Kumar Dogra, and TanujWala, “A Review Paper on Data Mining Techniques and Algorithms”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 5, May 2015.

[2] Jedrzejowicz, J., & Jedrzejowicz, P. (2020), “GEP-based classifier for mining imbalanced data. Expert Systems with Application”, 114058. doi:10.1016/j.eswa.2020.114058.

[3] Francesco Musumeci et al., “An Overview on Application of Machine Learning Techniques in Optical Networks”, arXiv:1803.07976v4 [cs.NI] 1 Dec 2018.

[4] Salaken, S. M., Khosravi, A., Nguyen, T., & Nahavandi, S. (2017), “Extreme learning machine based transfer learning algorithms: A survey”, Neurocomputing, 267, 516–524. doi:10.1016/j.neucom.2017.06.037.

[5] Miss. Mayuri S. Shelke,” A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique”, 2017 International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 03, Issue 04.

[6] https://link.springer.com/chapter/10.1007/978-3-642-24958-7_85

[7] https://www.diva-portal.org/smash/get/diva2:1116036/FULLTEXT01.pdf.

[8] https://dl.acm.org/doi/fullHtml/10.1145/3343440#Bib0105

[9] https://www.researchgate.net/publication/281817338_A_Survey_Data_Classification_Approaches_and_Tools/link/55f9620108aeba1d9f1bdaef/download

[10] http://www.ijarse.com/images/fullpdf/1521195654_Vedant612ijarse.pdf

[11] Ayon Dey [2016], "Machine Learning Algorithms: A Review ", International Journal of Computer Science and Information Technologies, Vol. 7 (3), 2016, 1174-1179.

[12] Sunpreet Kaur and Sonika Jindal [2016], “A Survey on Machine Learning Algorithms”, International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2763 Issue 11, Volume 3 (November 2016).

[13] Gao Huang, "Trends in extreme learning machines: A review", Department of Automation, Tsinghua University, Beijing 100084, China.

[14] Musatafa Abbas AbboodAlbadra and Sabrina Tiuna,” Extreme Learning Machine: A Review”, 2017 International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 12, pp. 4610-4623.

[15] http://www.ijesi.org/papers/RTSCA-2K17-2017/G4859.pdf.

[16] https://www.hindawi.com/journals/mpe/2019/3526539/

[17] Elreedy, D., & Atiya, A. F.,” A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for Handling Class Imbalance”, 2015 Information Sciences. Elsevier, Vol.502, pp. 32-63, DOI:10.1016/j.ins.2019.07.070.

[18] Wong, S. Y., Yap, K. S., & Li, X.,” A New Probabilistic Output Constrained Optimization Extreme Learning Machine”, 2020 IEEE Access, doi:10.1109/access.2020.2971012.

[19] R. Rustogi and A. Prasad, "Swift Imbalance Data Classification using SMOTE and Extreme Learning Machine," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862112.

[20] Z. Yuan and P. Zhao, "An Improved Ensemble Learning for Imbalanced Data Classification," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 408-411, doi: 10.1109/ITAIC.2019.8785887.

[21] D. Durga Prasad, D. V. Prasad, and K. N. Rao, "Imbalanced Data Using with-in Class Majority Under Sampling Approach," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-5, doi: 10.1109/ICECCT.2019.8869339.

[22] Lee, Z.-J., Lee, C.-Y., Chou, S.-T., Ma, W.-P., Ye, F., & Chen, Z., “A Distributed Intelligent Algorithm Applied to Imbalanced Data”, 2019 IEEE International Conference of Intelligent Applied Systems on Engineering (ICIASE), doi:10.1109/iciase45644.2019.9074009.

[23] M. Lv, Y. Ren and Y. Chen, "Research on imbalanced data: based on SMOTE-AdaBoost algorithm," 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE), Xiamen, China, 2019, pp. 1165-1170, doi: 10.1109/EITCE47263.2019.9094859.

[24] L. Xiang, G. Zhao, Q. Li, W. Hao, and F. Li, "TUMK-ELM: A Fast Unsupervised Heterogeneous Data Learning Approach," in IEEE Access, vol. 6, pp. 35305-35315, 2018, doi: 10.1109/ACCESS.2018.2847037.

[25] Y. Ge, D. Yue, and L. Chen, "Prediction of wind turbine blades icing based on MBK-SMOTE and random forest in imbalanced data set," 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, 2017, pp. 1-6, doi: 10.1109/EI2.2017.8245530.