A Quick Review on Classification and Clustering Methods in ML with Optimization Algorithms

K.Kalpanarani
School of Advanced Sciences
VIT University Chennai Campus, Chennai, Tamilnadu, India

G.Hannah Grace
School of Advanced Sciences
VIT University Chennai Campus, Chennai, Tamilnadu, India

Abstract- Machine learning is one of the widespread domain in the modern world. It deals with different kinds of data and mainly is the most encouraging sector of Big data. As good as all kinds of disciplines are available machine learning helps to resolve the complications and assist to make best pre-eminent results. Optimization is an iterative method. It compares many solutions until the best solution is reached. The decision-making process in machine learning will be done by optimization algorithms. In this paper we analyze different types of classification and clustering algorithms along with optimization techniques.

Keywords – Classification, Clustering, Optimization

I. INTRODUCTION

Machine learning is popular because of its quick adoption of trends and patterns. Human support is not needed once we developed the algorithm. By experience, we can improve the learning skills of the machine, all types of data and all dimensional can be examined during the dynamic background. Even though it has some flaws, accession of data, leads to inefficient methods, Time complexity of the algorithm, need good resources to implement the algorithms, the exploitation of results is a very big challenge and more error vulnerable. If we do not find the errors it will be continuous and affects the accuracy. To reduce that flaws many researchers have given different methods with the combination of more optimization methods. Classification and clustering with regards to structured and unstructured data is the major focus in machine learning. Before classifying or clustering the data undergoes pre-processing for more efficiency. We have many pre-processing methods in which the feature selection method [1] is more observant which is categorized into two category wrappers and filters. Wrappers use the existing classification algorithm in the search activity for the evaluation of the selected feature subset while the filters will select feature subset based on the performance evaluation metrics applied on the data. Filters are rapid than wrappers despite wrapper are efficient than filters. We have many classification algorithms [2] in which KNN, Naive Bayes, and decision tree are favoured and in clustering k-means, neural networks, Evolutionary computing, deep neural network, Agglomerative Hierarchical algorithm, DBSCAN, Expectation-Maximization (EM) are well-liked.

Optimization [3] is the branch of applied Mathematics. Optimization encompasses three main integrant which are objective function, constraints, and decisions. The definition of optimization is according to the problem description, to find the feasible solution of the problem we have to maximize or minimize the objective function by satisfying all the required constraints within the search space. The standard form of Mathematical optimization is

$$\text{Minimize } g_0(t) \qquad\qquad \text{Maximize } g_0(t)$$
$$\text{Subject to } g_i(t) \le d_i,\ I = 1, 2\ldots n \qquad\qquad \text{Subject to } g_i(t) \ge d_i,\ I = 1, 2 \ldots n$$

Here $t=(t_1, t_2 \ldots t_n)$ is the optimization variable of the problem, $g_0(x)$ is the objective function, $g_i(x)$, i= 1, 2 … m are the constraint function and both the objective function and constraint function are mapped from R^n to R and the constants $d_1, d_2 \ldots d_n$ are the bounds for the constraints. The vector x* is called the optimal solution of the problem.

The optimization algorithm is a process of iteratively comparing the numerous solution until the optimum solution is reached. There are many types of optimization problems. The optimization problems are constrained or unconstrained, problems can have none, one or more objective function and the variables can be discrete or continuous, some problems are static and some are dynamic and the decision variables can be integer, real-valued, deterministic or stochastic and the equations can be linear, nonlinear, quadratic or polynomial. Numerical methods of optimization [4] are Linear programming, Integer programming, Quadratic programming, Non-linear programming, Stochastic programming, Dynamic Programming, Combinatorial optimization, Constraint satisfaction etc. Advanced optimization methods are Hill Climbing, simulated annealing, Genetic algorithm, Ant Colony Optimization, etc. Optimization algorithms are employed to emphasize the efficiency of the classification and clustering algorithms. Optimization is a part of Machine Learning (ML). ML looks over the real-world problems and encoding into models by investigating and resolving the problem. ML needs the support of the optimization because of its best generalization, large problems scalability. The performance execution time and memory requirements are good, easy to implement, efficient manipulation of the model, fast convergence, handling complex problems with numerical stability. In ML the adaptability of an optimization plays an important responsibility. Like SVM, Bayesian Networks, Association rule, and DBN,we have different optimization techniques like swarm intelligence, bio-inspired algorithm, physics, and chemistry-based algorithms all of which contained in the nature-inspired algorithm. ML interplay with nature-inspired optimization algorithm to improvising stability, scalability, and accuracy. Some simulated nature-inspired algorithms [5] are Genetic algorithm (GA) [6], [7], Particle swarm optimization (PSO) [8], [9], Ant colony optimization (ACO) [10], [11], Whale optimization (WO) [12], [13], Bee colony optimization(BCO) [14], [15] , Bat optimization [16], [17], Grey Wolf Optimization (GWO) [18], [19], Cuckoo Optimization Algorithm (COA) [20], [21], Firefly Algorithm (FA) [22], Differential Evolution Algorithm (DE) [23], [24] etc.

The rest of this paper contains survey of literature of a combination of ML algorithm with optimization techniques and conclusion of the literature survey.

II. Literature Survey

The observation in paper [25] is how the techniques of data mining are manipulated to identify the prediction of the success rate of IVF (In-vitro Fertilization) treatment based on the high impact features. The proposed algorithm helps gynaecologists to suggest the patients for specific treatment and it is also helpful to increase the confidence level of the patients. Weka tool was used for evaluation. The IVF data sets were taken from the speciality test tube hospitals and IVF research centres in Tamilnadu for this experiment which contains 250 records and 27 features. In the data pre-processing feature selection have been made to remove redundant and irrelevant features which improves mining accuracy and reduces computational time and enhances the result. mRMR filter approach was used to select the high impact feature subset of four features among 27 features. Then the classification was carried out by the method of multilayer perceptron neural network. To know the performance of the proposed model the cross-validation was used to measure the confusion matrix and true pulse and false pulse which leads to the minimum number of high dominating features to predict the success rate of IVF treatment.

In the next study [26] the data set of liver disease patients were analysed. It is not easy to identify the people who are affected by liver disease. There may be people who are misled based on the symptoms. In this study, the authors have given the best opinion to overcome this problem. The data set taken contains of 583 patients and 11 attributes. The weka tool was used for classification. Primarily to analyse the data set, the pre-processing has taken place to remove the duplicates and filling the missing values then discretization was applied to convert the continuous data into categorical data. Afterwards, the following three methods Decision tree, Naive Bayes and NB Tree algorithm were used for classification. The accuracy and computational time was recorded for all the three methods. Optimizing the NB tree algorithm gives more accuracy than others even though the computational time of Naive Bayes was very less. Finally, they conclude that the NB tree algorithm gives the best liver disease classification.

AIDS is a very dangerous disease in the world if we predict this in earlier stages we can increase the life span of the affected patients. In this paper [27] the author proposed C4.5 classification algorithm to build a decision tree. The experimental study was done on HIV data set it was taken from NACO and government hospital in Sivakasi which contains 800 records and 9 attributes have been used. The R software was used for this experiment. Prior to classifying the data pre-processing is carried out for replacing the missing values by median values of the corresponding data. Then attribute selection was done using one-R and information gain by this only 6 attributes were selected for further classification. After that C4.5 Decision tree algorithm was used to classify the first stage of HIV or STD and other stages of HIV.

In the smart cities maximum of the industries and the customers unfavourable component is Transportation. To overcome this issue many of the researches invented many ideas but that did not have many economic benefits. By investigating many of the previous research and inspired by the economic problem and real-time data problem, in this paper [28] the researcher proposed a high-performance system for finding the correct location using lively transportation network and the total cost was minimized using the optimal transportation network. On the basis of the real-time big IoT data and GIS data, clustering centres were found using the k-means algorithm, before applying this DBN model was used to get the pre-classification results which were used as initial centres in the K-means to improve the clustering accuracy and computational performance. Traditional k-means does not converge faster so improved k-means was used to find the new centroid, where the weight factor helped to converge faster. To get the final optimal solution calculated minimum transportation cost between clustering centres and the supplier spots were used in the algorithm target index model. The proposed model applied in the hotel service locations within the out-ring of Tianjin city to get the best configuration centres. The basic data and geocoded data of 618 hotels were taken for this study which was observed from the internet. The DBN clustering algorithm were implemented 26 times and independently divided the service centres from 5 to 30 with in the region which was taken as initial centres in the improved k-means clustering in order to get the best amount of service centres with the minimal overall transportation and ordering cost. According to different transportation network and clustering centres, the annual transportation cost was calculated to get the optimal number of service centres with the lowest cost. To measure the performance of the proposed model the computational speed of the three methods namely Traditional K-Means, K-Means with transportation cost as weigh for new centroid and KM using DBN to pre-process data with transportation cost as weigh for new centroid according to the 5,15,20,25 service centres were compared. From these three methods, KM using DBN to pre-process with transportation cost as target index method computational speed was very faster than the other two. The optimal number of service centres implemented from the proposed model within out-Ring of Tianjin city is 25 according to the minimum total cost. The author suggested that the proposed model will give clear results when applying it to states, countries and continents. Further, when we applied this proposed model on other transportation business models the parameters of the model should be customized as per our requirement.

In the recent application sector, the data used were with multilevel learning and multi-label classification. According to the earliest studies of the multi-label feature selection, the greedy method was used to get the optimal subset selection. Here the score function was created which depends upon the dependency of the features. There were enormous subsets and the method was easily trapped into local optima which proves to be inefficiency. To enhance the efficiency of the multi-label classification the author [29] advocates the multi-label feature selection based on the numerical optimization. The score function was build based on mutual information and surviving numerical optimization to escape from the local optima. The different category of four datasets were used for experimental calculation and which was compared with existing unconventional multi-label classification FIMF, QPMLFS, MFS, D2F and PMU. The score function, search based evaluation and classification was evaluated with the multi-label K-nearest neighbour (MLKNN). The k value was set to be 10 and the number of feature selection as 30. Even with simplicity and excellence, it limited to time consumption and also it was not efficient all the time because of its first-order dependency.

The shortcomings of the cellular networks when increasing high data services and heavy distributed traffics are inefficient network capacity and the maximum coverage hole due to the imbalanced network configuration between the traffic demands and radio resources and the traditional methods are not suitable to give the best results for the dynamic changing wireless network. To fulfil the shortcomings the author [30] proposed a data mining approach, Weighted KNN based on a large-scale measurement statistics analysis to record the dynamic characteristics of radio resource demands and the value of “k” in KNN. The most accurate time series was determined. A greedy optimization technique was used for reconfiguration of the dynamic radio resource. To know the efficiency of the proposed study the statistical analysis was done over tens of thousands of cells for a duration of 15 months.

Efficient energy utilization is the major problem in the wireless Adhoc networks. The major pitfall is to control the unpredictability. To handle the uncertainty the fuzzy constraints applied clustering optimization technique was proposed [31]. The fuzzy constraints of the proposed model were, device energy, device position, device movement/speed and device hop-count which decided the active state of the network. By fuzzy constraints, the fitness value of each network node within the communication range of a maximum two-hop was calculated to select a cluster head. In case of uncertainty, new cluster head selection was taken and the network activities continued. To evaluate the new technique the network simulator, NS-2 was used. The size of the network was 1000 m * 1000 m and the range of the nodes of network varied from 50 to 250. The simulation was registered to 50 different runs for distinct nodes and clusters. The performance of the proposed technique had been compared with the methods

LEACH, MPO, EDC and TCACWCA, the metrics used for performance evaluation were no of clusters formed, network lifetime, average delay time and packet delivery ratio.

In distributed system accessing data at the correct time is predominant, it becomes difficult to generate a strong association rule. To generate robust association rule the author [32] proposed the W-tabular algorithm which is a combination of Weight assignment method and Quine-Mccluskey method. Weight assignment method removed the unessential data and Quine-Mccluskey method converted the pre-processing data into binary expression and joined the weights to get improved association of rule reduction rate and increased data processing time. This proposed model evaluated by the computer system of Intel ® core TM i53230M CPU@2.60GHz processing units and 6.00GB RAM with Windows XP in a distributed system by using a client-server model. The different transactional datasets like Mushroom, Accidents, Connect, Pumps and Chess were taken from FIMI Repository for experimental study to know the performance of the proposed model and the same datasets were used for comparing the proposed algorithm with apriori algorithm to know the well-built model.

The author [33] proposed a synthesis algorithm of TLBO and Fuzzy C-means with elicit property for data clustering (ETLBO-FCM). TLBO, a famous unconventional meta-heuristic algorithm which did not control any predominant parameters were the main bonus in this model. The Fuzzy C-means was popular because of its class overlapped property, this model was better than K-means and the elicit property was used to replace the duplicate solution with best solution throughout the population which decreases and leads to the duplicate solution. For the experiment 14 datasets were used in which 11 were taken from the UCI repository and 3 were artificially created, and it was done through MATLAB tool. To know the effectiveness and efficiency of the proposed model it was compared with K-means, GA –K means, PSO–K means, IPSO–K means, TLBO–K means, FCM, GA- FCM, PSO- FCM, IPSO-FCM, TLBO- FCM.

To minimize the procurement cost of the global cruise ship supply, the author [34] followed the multi-item joint ordering strategy. The objective function was combined to purchase, delivery and inventory based on the cruise distribution centre and proposed the improved fireworks algorithm (i.e.) the fireworks algorithm with inertia weight (WFWA), which speed up the convergence speed and not easily catch into the local optimum. The MATLAB software was used to the experimental analysis. To know the performance of the proposed model, it was compared with the existing models FWA, GA and PSO, for more accuracy each model ran 20 times, and the iterations were 500.

To enhance the accuracy of the classification the author [35] suggested an innovative shuffling based coaction of optimization clustering algorithm which was COA-GA. It was applied in the cancer classification to get to know the more supreme genes to classify using the shuffling method. The four microarray data sets were used taken from public sectors and broad institute. Data Pre-Processing was done to remove the error data to be used for clustering. The different types of optimization techniques like K-means, C-means, Hierarchical, GA, PSO, COA, COA-GA, and no cluster were used to select the gene selection, by which the most dominant genes only selected for further evaluation. Using SVM or MLP the genes were classified. The performance measures sensitivity, specificity, and accuracy were used to estimate the proposed algorithm. When compared with other optimization methods the COA-GA outperforms and also it reaches the optimum level in the minimum number of iterations, and SVM was given best when compared with MLP, and for more accuracy, it was running 100 times.

We have many intelligence optimization algorithm to solve the many problems. The author implemented the Sine Cosine algorithm with multigroup and multistrategy to conclude the optimization solution of the problems. To realize more robustness the author [36] introduced multigroup and multistrategy in the existing SCA algorithm. In the multigroup, each population had the same number of individuals, and the multistrategy had a different update strategy, which is divided into two types. , First rand strategy, used to enhance the global optimization, and the second-best strategy, used to stable the solution of the problem and convergence speed, and the population was divided into two groups as odd and even and the two strategies were applied on it to get the best solution. The programming tool MATLAB was used to get the experimental results. To test the performance of the MMSCA model, 19 standard benchmark functions were used which contains unimodal, multimodal and composite functions and also it was compared with the existing algorithm PPSO, PSO, DE, and each algorithm ran 30 times, a maximum number of iterations was 1000 and each population size was set to 30. The proposed algorithm MMSCA was applied in the real-time problem CVRP, in which the number of population in each group is 50, the algorithm ran 30 times and the number of iterations was 10000 and also it was compared with PSO and GA to know the practicability of the MMSCA algorithm.

Document clustering is a big problem in the field of text mining and information retrieval. To enhance the clustering accuracy the author [37] proposed the spectral clustering algorithm with particle swarm optimization (SCPSO). Before using the algorithm pre-processing was done by tokenization, stop word removal and stemming, then from the document Eigenvectors of the weighted undirected graph was derived by the spectral clustering algorithm.

Particle swarm optimization was used to find the optimal cluster and check which gives the best result for high volume documents. The standard datasets like Reuters, 20 Newsgroups, and TDT2 were used to test the proposed model. To know the performance of SCPSO which was compared with EM, SK-Means, PSO and the performance measures were Accuracy, NMI, ARI.

To give a more powerful and well-organized method of data clustering, the author [38] proposed the particle swarm optimization with selective particle regeneration (SRPSO), in which an unbalanced parameter setting and particle regeneration operation was applied in the basic particle swarm optimization. An unbalanced parameter setting was permitted for the fast convergence but it was easily trapped into local optima. To run off from that issue the particle regeneration was used, in which the particle that fell into the local optima was reassigned by the new position. The hybrid algorithm was proposed which was a combination of K-means and SRPSO, in which the initial particles are chosen by K-means and then SRPSO applied. The MATLAB tool was used to evaluate the proposed methods, for experiment seven benchmark data sets, and two synthetic datasets were used. To know the outcome of proposed models SRPSO and KSRPSO, it was compared with PSO, KNM-PSO, KGA, PSO+R1, PSO+R2.

According to the shortcomings of the various combination of PSO methods, the author [39] proposed the PSO algorithm with kernel density estimation (KDE). This filled the imbalance between exploitation and exploration. In the proposed algorithm first, the particles are initialized by uniform distribution with the semi-random process which was useful to escape from the initial difficulties, secondly, the KDE was used to find the best position of the particle, which ensured the non-early convergence and finally the multi-dimensional gravitational learning coefficient was estimated by a universal gravity rule. Then the velocity of the particle was updated and the optimal solution found, and the particle moved in all directions and distributed unequally in feature space. The different kinds of experiments were done to test the proposed algorithm where MATLAB tool was used and the 11 standard datasets were taken from the UCI repository. In the beginning, the comparison was done with variations in the proposed algorithm with HPSO by normalized and un-normalized datasets, the comparison results contained the values of Friedman Aligned Rank, Holm APV, and p-value. Then the proposed method was compared with the existing methods PSO, PSO-K-Means, K-means++, FCM, PSO-KHM, in which Classification Accuracy, Standard deviation, and Dunn Index were measured. The running time also measured, which given the best result when compared to other methods. It consumed a shorter time only when applied on high-dimensional dataset.

In the modern world transportation traffic is one of the puzzling criteria. In addition to enhance an intelligent traffic control system, the author [40] proposed ant colony optimization with the internet of vehicles (IoV) in which the vehicle was dynamically communicated. In this study IoV data of the Vellore district, Tamilnadu, India, was used. In the proposed method first, the street map was segmented into a small number of distinct maps and then the ant colony optimization method was used in each map to get the optimal routes and to update the traffic intensity and pheromone fuzzy logic was used. To evaluate the proposed method it was compared with the existing shortest path methods Dijkstra algorithm, Kruskal's algorithm, Prim's algorithm.

In the IoV data, the interaction between vehicles and the control of the network is very complicated tasks in the fast-moving vehicles and the quickness of data delivery. The author [41] learned that if there is a need to increase the efficiency of the method we need to regulate parameters and suggested to reconstruct ant colony optimization for IoV for clustering (CACOIOV), which contained two stages, At the first stage this method chooses the initial node intelligently in the search space which increases the reliability of the interactions between the network and at the second stage introduced unfamiliar method to increase the convergence speed. One more method was proposed together with CACOIOV was dynamic aware transmission range on local traffic density (DA-TRLD), which sustained the network connectivity. The evaluation of the method was handed over through NS-2 simulations and compared with ACO and AODV with regarded packet delivery ratio, drop ratio, delay, cluster number, throughput.

Decision support systems are used to analyze huge data from raw data, documents, personal knowledge, and business models, and give useful information. Based on the addition of data in the relational database, we have to reconstruct and reorganize the enormous and unordered data based on the decision. To get the optimal knowledge from the database of the decision support system the author [42] expressed the coaction of information entropy and ant colony optimization. This proposed model applied to the real database of website construction in the e-commerce field. The path analysis was done. The performance of the proposed method was recorded and it was compared with the traditional method on the basis of misclassification error rate and number of iterations.

Identifying communities in social networks is challenging and important in the world. We have many methods to solve the issues but for more efficiency, the author [43] proposed a new model. For this analysis, they have taken Twitter self-centered social network dataset which contains 2000 users from the SNAP website. They have executed and evaluated the proposed model using the MATLAB R2016a programming language. The similarity matrix of the features communication topology and tweets between users using the different similarity methods is calculated. The initial clusters using the different type of clustering algorithm are formed using matrix and finally using cuckoo optimization technique have been used to create final clusters. The clusters were evolved due to the integration of several small clusters. The performance of the proposed model was compared with the existing models CC-GA and MDCL, they have measured by metric, modularity, and silhouette criterion.

We have many data clustering methods but some methods failed for multi-dimensional data. The author [44] proposed a fuzzy cuckoo optimization algorithm to overcome this problem. MATLAB tool was used for evaluation. First, the algorithm gives random solutions to the large data sets, and for each solution cost value had been calculated and then fuzzy logic is applied to get the optimal solution. The proposed algorithm was used in the seven benchmark datasets and also it was compared with Blackhole, CS, K-Means, GSA and PSO to know the better performance. For better result, it was executed 10 times for 200 iterations.

For data clustering, we have many methods but many of them are sensitive to the initial clustering center selection and uselessness to deal with noisy, outlier, and imbalanced data. To overcome this issue the author [45] proposed Magnetic optimization algorithm clustering (MOAC), where data points are applied to magnetic force directly to the centroid particles to update the position to get the best centroid particle for clustering. When the force value is reached to zero it had been concluded that the optimum position was obtained. The MATLAB tool was used for experimental analysis. They have used twelve benchmark data sets for experimental study. Out of which, eleven data sets taken from UCI repository and one synthetic data set aggregation. To know the performance of the proposed algorithm MOAC, it was compared with HYBRID_DE, MIN_MAX, PSO, K-Means, KFCM and the evaluation factors, Normalized mutual information(NMI), Rand index(RI), Purity, Accuracy were used, Each algorithm ran 30 times and the maximum iteration was set to 200. The drawback of this MOAC need prior knowledge of the number of Clusters.

The main drawback of the K-means algorithm is it easily traps into local optimization which is accompanied by inefficient and incorrect initialization. This stimulated the author [46] to come up with Electromagnetic optimization based clustering (ELMC) to overcome the drawback and it was a modified method of electromagnetic field optimization (EFO). In the ELMC the attraction-repulsion was deployed to avoid the initial centroid difficulties. The datasets were taken from the UCI repository and the MATLAB tool was handled for the statistical experiments. The execution of the ELMC was compared with other meta-heuristic methods K-Means, MKFC, KFC, PCM, ACO by manipulated the measures NMI, RI, Purity.

The benefits of feature selection decreases overfitting with training time and increases accuracy. To enhance this benefits the author [47] implemented the unconventional method of hybridization whale optimization algorithm (WOA) with simulated annealing (SA), which involves two hybridization model that was low-level teamwork hybrid (LTH). Here the SA was placed as a local operator in WOA, this was denoted as WOASA-1 and high-level relay hybrid (HRH), in this SA was applied to the best solution after the completion of all iterations in WOA, denoted as WOASA-2. To reduce the selection of not robust solution tournament selection mechanism was used in the proposed methods which was denoted as WOASAT-1 and WOASAT-2. For the experimental study, the MATLAB tool was used and 18 benchmark datasets were taken from the UCI repository. The evaluation and comparison was done based on the classification accuracy and average selection size on the four proposed models and WOA, the outperformed model was WOASAT-2, which was compared with ALO, PSO, GA on the basis of the mean, best and worst fitness value, in this also WOASAT-2 was balanced search and utilization.

Internet and big data are the fast-growing fields which contains high-dimensional and large-scale data. When applying conventional feature selection on this it leads to inefficient and less classification accuracy. To get rid of this the author [48] proposed the synthesis of WOA and parallel computing model of the spark environment, which disseminate data across the cluster and process the data in parallel, this approach enhanced the efficiency of the model. The experiment of the proposed model was applied on the 4 datasets grasped from UCI repository and it was compared with GA, PSO and WOA on the support of run time and speedup ratio of the feature selection.

Previously many of the researchers implemented the effective unconventional feature selection methods for the continuous space search and it may or may not work well for the binary space search from the inspiration of this the author [49] invented synergist of Grey Wolf optimization and particle spam optimization of the binary type (BGWOPSO) for the robust feature selection. The wrapper method of the KNN classifier with the Euclidean separation metric was involved to find the best solution. The experiment was done on the 18 standard benchmark datasets which contain small, medium and large data size and concluded that the proposed model gave the best results when compared with other hybrid models BGWO, BPSO, BGA, WOASAT-2 regarded of the accuracy, an average number of selected features and mean, best, worst fitness values. The proposed method controls the difference between examination and utilization during the optimization iterations.

Many high dimensional and features of data are extremely surplus in the real time data. Dimensionality reduction only won’t give the best result but we need to concentrate on the data reduction too. The author [50] suggested an innovative feature selection method based on artificial Bee colony optimization and gradient boosting (ABCoGB).

Primarily the unnecessary features were diminished using the artificial bee colony optimization and then the resultant dataset involved in the gradient boosting decision tree to get the final best selection of features. The proposed model implemented through python 3.5 software and applied on 8 benchmark datasets from the UCI repository. According to the classification accuracy, the suggested method compared with GBDT, VT, SKB, RFE, L12 and PCA to know the excellence.

When dealing with big data we concentrate on execution time and the difficulties faced by the use of large data. The traditional methods are not applicable for the large data so the author [51] proposed artificial bee colony optimization (ABC) method to minimize the execution time for large data and to optimize the clusters for all sizes of data. The ABC method implemented on a single node and multi-node Hadoop environment using mapper and reducer stages. In the mapper stage best fitness value was found by copying the use of bee behavior and in reducer stage, the cluster was optimized using probability value based on the copying of involved and observer bee behavior. The probability value was used to find the classification error. The glass benchmark data was taken from the UCI repository which contains 10 attributes for the experiment. The evaluation was compared with ABC, PSO and DE concerned with execution time and classification error. Finally decided that multi-node Hadoop environment given the best result.

Most of the nature-inspired algorithms performed excellently in the low-dimensionality data. Come into the high dimensional data need some improvement on the conventional algorithm. The author [52] proposed two methods PCA-BA and PCA-LBA which is a combination of principal component analysis (PCA), bat algorithm (BA), and golden section (GS) method. In the golden section (GS) method the two parameters correlation threshold and generation threshold were initialized to enhance the performance of the proposed methods. The experiment was done on the CEC'2008 benchmark dataset which contains two unimodal functions and five multimodal functions and implemented with MATLAB tool. The performance of the GSPCA-LBA was compared with PSO, TVAC, CS, CPCA-BA, CPCA-LBA and GSPCA-BA based on the different dimensions and using Friedman test and Wilcoxon test which concluded that GSPCA-LBA outshined.

Internet of things is the most popular field in which we consider wireless sensor big data sensing system which consumes more energy. Low energy adaptive clustering hierarchy (LEACH) method reduces energy consumption even though by picking a random cluster center it has a chance to get a cluster center far from the base station. To overcome this issue the author [53] established modified bat algorithm with weighted harmonic centroid (WHC) which was used to improve the local optimization search and to know the performance of WHC five more centroid methods arithmetic centroid, geometric centroid, harmonic centroid, weighted arithmetic centroid, weighted geometric centroid were developed and compared. The reflection of the results among these methods showed that WHCBA was supreme. Besides the combination of WHCBA with LEACH method excelled when compared to LEACH and other algorithms based on the Friedman and Wilcoxon tests. The data was collected from CEC 2013 which contained 28 benchmark function for experimental analysis.

The author [54] proposed an adaptive cat swarm optimization (ACSO), which is a combination of CSO and APSO. Some strategies were followed in the proposed model to improve the convergence speed and search capabilities, which are, in the tracing mode the adaptive adjustment to its parameters and memory factor 'y' was introduced and the radius range was added in the search position. The MATLAB software was used for the experimental study. The experiment of the proposed model done in the 23 benchmark functions which contained unimodal, multimodal, fixed-dimension multimodal functions and it was compared with the existing methods PSO, APSO and CSO to know the performance of the ACSO based on the average and standard deviation of the methods and each algorithm ran 10 times, the population size of each algorithm was 100 and the number of iterations was 500. The ACSO was applied in the vehicle routing problem (VRP) to know the practicability of the method. The iteration of the method was 300 times and it was applied to the different cases like different numbers of customers and vehicles. It was compared with the methods like PSO, APSO and CSO based on the best cost and time.

III.CONCLUSION

In this paper, we have reviewed some literature and learnt from this review that the combination of the machine learning and optimization techniques outperforms when compared with other trivial machine learning algorithms and conventional optimization algorithms. We cannot conclude that the combination of ML and optimization will not fix the problems of all disciplines. Despite the novel methods that give the best results for that specific cases only. And so we need attention for all the cases and some of the methods can only be fit with small scale problems. When we use the same method for large scale problems it may not be effective and efficient so we have to concentrate on scalability when we creating the models. We are facing obstacles in the theoretical interpretation because we know how it’s working empirically and do not know the conditions prevailing. When we calibrate the parameters it will affect the performance though we have to find how to adjust and adjusting value margin. Mathematical framework should be done for more insights of stability, convergence, rate of convergence, and robustness. We are applying most of the novel methods on benchmark sets only not on the real-time problems , sometimes it was not worked well for the real-time problems nevertheless it was worked well on the benchmark sets, so need more attention when we choosing the benchmark sets. The majority of the methods use the performance measures to evaluate the performances like accuracy, computational time, stability and success rate, for reasonable and precise comparison we need unique performance measure structure. The resources for the implementation of the algorithms are restricted and we have to consider all the resources when we model the machine. These learnings will be useful for furthermore research.

REFERENCES

[1] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, “A review of feature selection methods on synthetic data,” *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, Mar. 2013, doi: 10.1007/s10115-012-0487-8. A. A. Reddy and B. N. Chatterji, "A new wavelet based logo-watermarking scheme," Pattern Recognition Letters, vol. 26, pp. 1019-1027, 2005.

[2] N. Jothi, N. A. Rashid, and W. Husain, “Data Mining in Healthcare – A Review,” *Procedia Comput. Sci.*, vol. 72, pp. 306–313, 2015, doi: 10.1016/j.procs.2015.12.145.

[3] Boyd, S., Boyd, S. P., and Vandenberghe, L., *Convex Optimization*. Cambridge university press, 2004.

[4] Rustagi , J.S., *Optimization techniques in statistics*. Elsevier, 2014.

[5] Yang, X. S., Nature-inspired optimization algorithms. Elsevier, 2014.

[6] D. Whitley, “A genetic algorithm tutorial,” *Stat. Comput.*, vol. 4, no. 2, Jun. 1994, doi: 10.1007/BF00175354.

[7] U. Maulik and S. Bandyopadhyay, “Genetic algorithm-based clustering technique,” *Pattern Recognit.*, vol. 33, no. 9, pp. 1455–1465, Sep. 2000, doi: 10.1016/S0031-3203(99)00137-5.

[8] Clerc, M., *Particle swarm optimization*, vol. 93. John Wiley & Sons., 2010.

[9] Yuhui Shi, “Particle swarm optimization: developments, applications and resources,” in *Proceedings of the 2001 Congress on Evolutionary Computation*, May 2001, vol. 1, pp. 81–86, doi: 10.1109/CEC.2001.934374.

[10] M. Dorigo and G. Di Caro, “Ant colony optimization: a new meta-heuristic,” in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, Washington, DC, USA, 1999, pp. 1470–1477, doi: 10.1109/CEC.1999.782657.

[11] C. Blum, “Ant colony optimization: Introduction and recent trends,” *Phys. Life Rev.*, vol. 2, no. 4, pp. 353–373, Dec. 2005, doi: 10.1016/j.plrev.2005.10.001.

[12] M. M. Mafarja and S. Mirjalili, “Hybrid Whale Optimization Algorithm with simulated annealing for feature selection,” *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017, doi: 10.1016/j.neucom.2017.04.053.

[13] S. Mirjalili and A. Lewis, “The Whale Optimization Algorithm,” *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[14] D. Teodorović, “Bee Colony Optimization (BCO),” in *Innovations in Swarm Intelligence*, vol. 248, C. P. Lim, L. C. Jain, and S. Dehuri, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 39–60.

[15] D. Teodorovic, P. Lucic, G. Markovic, and M. D. Orco, “Bee Colony Optimization: Principles and Applications,” in *2006 8th Seminar on Neural Network Applications in Electrical Engineering*, Belgrade, Serbia & Montenegro, 2006, pp. 151–156, doi: 10.1109/NEUREL.2006.341200.

[16] A. H. Gandomi, X.-S. Yang, A. H. Alavi, and S. Talatahari, “Bat algorithm for constrained optimization tasks,” *Neural Comput. Appl.*, vol. 22, no. 6, pp. 1239–1255, May 2013, doi: 10.1007/s00521-012-1028-9.

[17] A. H. Gandomi and Amir Hossein Gandomi, “Bat Algorithm: A Novel Approach for Global Engineering Optimization,” *Eng. Comput.*, vol. 25, no. 5, pp. 464–483, 2012, doi: https://doi.org/10.1108/02644401211235834.

[18] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey Wolf Optimizer,” *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.

[19] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, “Grey wolf optimizer: a review of recent variants and applications,” *Neural Comput. Appl.*, vol. 30, no. 2, pp. 413–435, Jul. 2018, doi: 10.1007/s00521-017-3272-5.

[20] R. Rajabioun, “Cuckoo Optimization Algorithm,” *Appl. Soft Comput.*, vol. 11, no. 8, pp. 5508–5518, Dec. 2011, doi: 10.1016/j.asoc.2011.05.008.

[21] H. Kahramanli, “A Modified Cuckoo Optimization Algorithm for Engineering Optimization,” *Int. J. Future Comput. Commun.*, pp. 199–201, 2012, doi: 10.7763/IJFCC.2012.V1.52.

[22] X.-S. Yang and X. He, “Firefly Algorithm: Recent Advances and Applications,” *Int. J. Swarm Intell.*, vol. 1, no. 1, p. 36, 2013, doi: 10.1504/IJSI.2013.055801.

[23] A. K. Qin, V. L. Huang, and P. N. Suganthan, “Differential Evolution Algorithm With Strategy Adaptation for Global Numerical Optimization,” *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 398–417, Apr. 2009, doi: 10.1109/TEVC.2008.927706.

[24] A. K. Qin and P. N. Suganthan, “Self-adaptive Differential Evolution Algorithm for Numerical Optimization,” in *2005 IEEE Congress on Evolutionary Computation*, Edinburgh, Scotland, UK, 2005, vol. 2, pp. 1785–1791, doi: 10.1109/CEC.2005.1554904.

[25] M. Durairaj and R. N. Kumar, “Data Mining Application on IVF Data For The Selection of Influential Parameters on Fertility,” vol. 2, no. 6, p. 6

[26] S. N. N. Alfisahrin and T. Mantoro, "Data Mining Techniques for Optimization of Liver Disease Classification," in *2013 International Conference on Advanced Computer Science Applications and Technologies*, Kuching, Malaysia, Dec. 2013, pp. 379–384, doi: 10.1109/ACSAT.2013.81.

[27] G. S. Lakshmi and P. I. Devi, "Prediction of anti-retro viral for HIV and STD patients using data mining technique," in *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, Srivilliputhur, Mar. 2017, pp. 1–6, doi: 10.1109/ITCOSP.2017.8303084.

[28] J. Yang, Y. Han, Y. Wang, B. Jiang, Z. Lv, and H. Song, "Optimization of real-time traffic network assignment based on IoT data using DBN and clustering model in smart city," *Future Gener. Comput. Syst.*, vol. 108, pp. 976–986, Jul. 2020, doi: 10.1016/j.future.2017.12.012.

[29] H. Lim, J. Lee, and D.-W. Kim, "Optimization approach for feature selection in multi-label classification," *Pattern Recognit. Lett.*, vol. 89, pp. 25–30, Apr. 2017, doi: 10.1016/j.patrec.2017.02.004.

[30] Z. Feng, X. Li, Q. Zhang, and W. Li, "Proactive Radio Resource Optimization With Margin Prediction: A Data Mining Approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9050–9060, Oct. 2017, doi: 10.1109/TVT.2017.2709622.

[31] A. S. Mohammed, S. B. B, S. B. M. S, A. P. N, and V. K, "FCO — Fuzzy constraints applied Cluster Optimization technique for Wireless AdHoc Networks," *Comput. Commun.*, vol. 154, pp. 501–508, Mar. 2020, doi: 10.1016/j.comcom.2020.02.079.

[32] B. Waghamare and Y. Bodhe, "Data mining technique for reduction of association rules in distributed system," in *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, Pune, India, Sep. 2016, pp. 415–418, doi: 10.1109/ICACDOT.2016.7877619.

[33] J. Nayak, B. Naik, D. P. Kanungo, and H. S. Behera, "A hybrid elicit teaching learning based optimization with fuzzy c-means (ETLBO-FCM) algorithm for data clustering," *Ain Shams Eng. J.*, vol. 9, no. 3, pp. 379–393, Sep. 2018, doi: 10.1016/j.asej.2016.01.010.

[34] L. Huang and J. Yang, "An Improved Swarm Intelligence Algorithm for Multi-Item Joint Ordering Strategy of Cruise Ship Supply," *Math. Probl. Eng.*, vol. 2020, pp. 1–9, Mar. 2020, doi: 10.1155/2020/5048629.

[35] V. Elyasigomari, M. S. Mirjafari, H. R. C. Screen, and M. H. Shaheed, "Cancer classification using a novel gene selection approach by means of shuffling based on data clustering with optimization," *Appl. Soft Comput.*, vol. 35, pp. 43–51, Oct. 2015, doi: 10.1016/j.asoc.2015.06.015.

[36] Q. Yang, S.-C. Chu, J.-S. Pan, and C.-M. Chen, "Sine Cosine Algorithm with Multigroup and Multistrategy for Solving CVRP," *Math. Probl. Eng.*, vol. 2020, pp. 1–10, Mar. 2020, doi: 10.1155/2020/8184254.

[37] R. Janani and S. Vijayarani, "Text document clustering using Spectral Clustering algorithm with Particle Swarm Optimization," *Expert Syst. Appl.*, vol. 134, pp. 192–200, Nov. 2019, doi: 10.1016/j.eswa.2019.05.030.

[38] C.-Y. Tsai and I.-W. Kao, "Particle swarm optimization with selective particle regeneration for data clustering," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6565–6576, Jun. 2011, doi: 10.1016/j.eswa.2010.11.082.

[39] M. Alswaitti, M. Albughdadi, and N. A. M. Isa, "Density-based particle swarm optimization algorithm for data clustering," *Expert Syst. Appl.*, vol. 91, pp. 170–186, Jan. 2018, doi: 10.1016/j.eswa.2017.08.050.

[40] P. M. Kumar, U. Devi G, G. Manogaran, R. Sundarasekar, N. Chilamkurti, and R. Varatharajan, "Ant colony optimization algorithm with Internet of Vehicles for intelligent traffic control system," *Comput. Netw.*, vol. 144, pp. 154–162, Oct. 2018, doi: 10.1016/j.comnet.2018.07.001.

[41] S. Ebadinezhad, Z. Dereboylu, and E. Ever, "Clustering-Based Modified Ant Colony Optimizer for Internet of Vehicles (CACOIOV)," *Sustainability*, vol. 11, no. 9, p. 2624, May 2019, doi: 10.3390/su11092624.

[42] Y. Peng, X. Yang, and W. Xu, "Optimization Research of Decision Support System Based on Data Mining Algorithm," *Wirel. Pers. Commun.*, vol. 102, no. 4, pp. 2913–2925, Oct. 2018, doi: 10.1007/s11277-018-5315-3.

[43] A. Rezaeipanah, M. Mojarad, and S. M. F. Hosseini, "Using Cuckoo Optimization Algorithm to Identify Communities in Social Networks," vol. 6, p. 8, 2019.

[44] E. Amiri and S. Mahmoudi, "Efficient protocol for data clustering by fuzzy Cuckoo Optimization Algorithm," *Appl. Soft Comput.*, vol. 41, pp. 15–21, Apr. 2016, doi: 10.1016/j.asoc.2015.12.008.

[45] N. Kushwaha, M. Pant, S. Kant, and V. K. Jain, "Magnetic optimization algorithm for data clustering," *Pattern Recognit. Lett.*, vol. 115, pp. 59–65, Nov. 2018, doi: 10.1016/j.patrec.2017.10.031.

[46] N. Kushwaha, M. Pant, and S. Sharma, "Electromagnetic optimization-based clustering algorithm," *Expert Syst.*, Dec. 2019, doi: 10.1111/exsy.12491.

[47] M. M. Mafarja and S. Mirjalili, "Hybrid Whale Optimization Algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017, doi: 10.1016/j.neucom.2017.04.053.

[48] Chen, H. et al., "A Spark-based Distributed Whale Optimization Algorithm for Feature Selection," in In 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Sep. 2019, vol. 1, pp. 70–74.

[49] Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection," *IEEE Access*, vol. 7, pp. 39496–39508, 2019, doi: 10.1109/ACCESS.2019.2906757.

[50] H. Rao *et al.*, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput.*, vol. 74, pp. 634–642, Jan. 2019, doi: 10.1016/j.asoc.2018.10.036.

[51] S. S. Ilango, S. Vimal, M. Kaliappan, and P. Subbulakshmi, "Optimization using Artificial Bee Colony based clustering approach for big data," *Clust. Comput.*, vol. 22, no. S5, pp. 12169–12177, Sep. 2019, doi: 10.1007/s10586-017-1571-3.

[52] Z. Cui, F. Li, and W. Zhang, "Bat algorithm with principal component analysis," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 3, pp. 603–622, Mar. 2019, doi: 10.1007/s13042-018-0888-4.

[53] Z. Cui, Y. Cao, X. Cai, J. Cai, and J. Chen, "Optimal LEACH protocol with modified bat algorithm for big data sensing systems in Internet of Things," *J. Parallel Distrib. Comput.*, vol. 132, pp. 217–229, Oct. 2019, doi: 10.1016/j.jpdc.2017.12.014.

[54] X.-F. Ji, J.-S. Pan, S.-C. Chu, P. Hu, Q.-W. Chai, and P. Zhang, "Adaptive Cat Swarm Optimization Algorithm and Its Applications in Vehicle Routing Problems," *Math. Probl. Eng.*, vol. 2020, pp. 1–14, Apr. 2020, doi: 10.1155/2020/1291526.