

ALGORITHMS OF MACHINE LEARNING

Mrs.Sangeeta Yadav

Faculty Head (Computer Science & Applications)

S.M.Joshi College Hadapsar Pune-28

Guide Name- Dr.Sanjaysingh Bharodia *Abdul Kalam University Indore*

ABSTRACT

MLA are described in this work. These algorithms are utilised for a broad variety of uses, including data mining, image processing, predictive analytics, and so on. Furthermore, deep learning, which is a subset of machine learning, can intelligently examine data on a huge scale. We provide a detailed overview of several machine learning methods in this article, which may be used to improve an application's intelligence and capabilities. The digital world is awash with data, including (IoT) data, cybersecurity data, mobile data, corporate data, social media data, health data, etc. in the present Fourth Industrial Revolution.

KEYWORDS

Machine Learning, Deep Learning, Artificial Intelligence, Data Science, Data-Driven Decision Making, Predictive Analytics, Intelligent Applications.

1. INTRODUCTION

There has been a lot of study in the field of machine learning to determine the applicability of algorithms in a domain for a particular specific goal, such as a classification or regression issue. The applicability of performance indicators for various areas has also been thoroughly investigated. Precision and recall are commonly used in information retrieval applications. The (AUC) is preferred by the medical community. Regardless of domain, accuracy is still a major criterion for evaluating the performance of ML models.

It's not unusual for a learning model to perform well on one parameter while failing to perform well on another for the same job. This section of the study compares (MLAs) on data from various IoT applications, categorising the data as binary or multiclass.

1.1. METHODOLOGY

The approach for comparing MLAs, as well as its statistical validation, is shown in Figure 1.

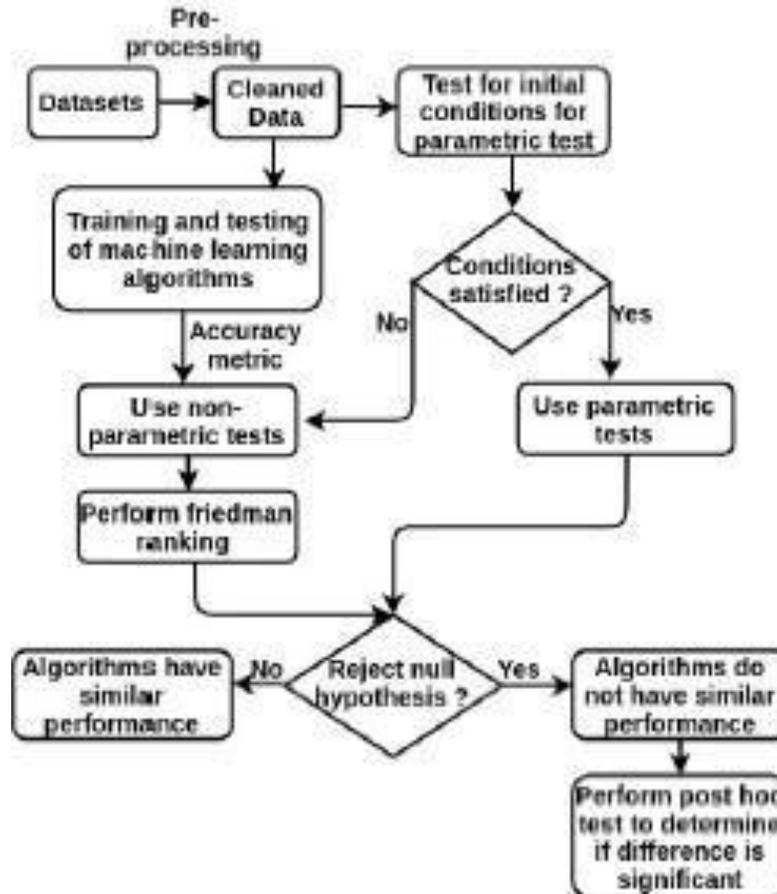


FIGURE 1: METHODOLOGY

Each dataset in figure 1 is cleaned for missing values or values that are out of range. Individual datasets are used to train and evaluate each learning algorithm. On each data set, the accuracy values for each algorithm are recorded. Data sets are checked for qualities to be satisfied for conducting parametric tests in order to validate experimental results. Parametric tests are used to statistically validate results if the qualities of normality, independence, and homoscedasticity are met. If any of the aforementioned characteristics are not met, non-parametric tests are used.

After completing a non-parametric test, a judgement may be made regarding whether to accept Take a position on the null hypothesis or reject it. Whether the hypothesis is rejected, conducted to see if the difference in algorithm performance is real.

1.2. MACHINE LEARNING ALGORITHMS

This study examines more than 30 papers in which machine learning techniques are employed for data analytics categorization or regression tasks. The commonly used machine learning methods for smart

data analysis and their online applications. They are divided into groups whose family they belong to. In our study, we employed the same techniques for our comparative analysis.

TABLE 1.
ML ALGORITHMS FOR ONLINE DATA ANALYTICS

ML ALGORITHMS FAMILIES	PROCESSING TASKS	USE CASES	OPTIMIZING METRIC
BAYESIAN			
Naïve Bayes (NB)	Classification	Agriculture, citizenry, and health care are all examples of smart technologies.	Food safety, Disease prediction
Discriminate quadratic analysis (QDA)	Classification / regression	Monitoring air pollution	Air Quality Predictive
Gaussian Process (GP)	Regression/classification	Smart traffic	Traffic predictions
Instance based			
Support Vector Machine (SVM)	Classification	Smart Trafic, Health	Classification of data
K-Nearest Neighbors (KNN)	Classification	Smart Citizens	Passenger travel pattern
Regression			
Linear Regression (LR)	Classification/Regression	Weather forecasting	Weather prediction
Decision tree			
Decision Tree (DT)	Regression/classification	Smart Grid for Weather	Air Quality Predictive
Neural network			
(MLP)	Classification/Regression/clustering	Reducing energy consumption	Reducing Energy consumption

Combining models (Ensemble models)			
Random Forest (RF)	Regression/classification	Agriculture and health care are both smart ideas.	Predictions of yield and disease
Boosting the gradient (GB)	Regression/classification	Parking Smart	Enhancing parking accessibility
AdaBoost (AB)	Regression/classification	Smart parking	Enhancing parking accessibility

1.3. CROSS VALIDATION

Different cross validation strategies are used to minimise biased learning. Three validation techniques are used to assess algorithm performance: (HoV), 10-fold cross validation (10FCV), and 5*2 cross validation (5*2CV). These validation procedures regulate aspects like as training set selection and test set selection, which have a significant impact on MLA performance. There is a lot of instability in these techniques of validation, but they are still neutral.

On all datasets, algorithms are run using each validation approach, and assessment is done using the most generally used classification job evaluation measure, "accuracy." Each validation approach is carried out in such a way that a total of 30 result samples are obtained. The final value for "accuracy" for each dataset is then calculated using an independent algorithm. Accuracy is a frequently used performance parameter in many machine learning applications. The most critical parameter in security, smart lighting, smart transportation, and seismic hazard applications is accuracy.

1.4. BONFERRONI-DUNN TEST

If the disparities in their average rankings across all datasets are more than or equal to the value of (CD), the performance of two algorithms is considerably different . Equation 3.4 is used to compute (CD):

$$CD = q_{\alpha} \sqrt{\frac{(K+1)}{6N}} \quad (3.4)$$

where $q_{\alpha} = 2.773$ is critical values for bonferroni dunn test which includes control algorithm. K is the number of algorithms and N is the number of datasets.

The bonferroni-dunn graph for binary class data is shown in Figure 2, with the bar height representing the rank of the algorithm determined using friedman ranking for the associated validation approach. The algorithm whose bar is higher than the threshold value the sum of the rank of the best performing algorithm,

the control, and (CD)) performs poorly compared to the control algorithm. For each form of validation, the horizontal line in a graph represents a threshold value, which is equal to the average rank plus (CD).

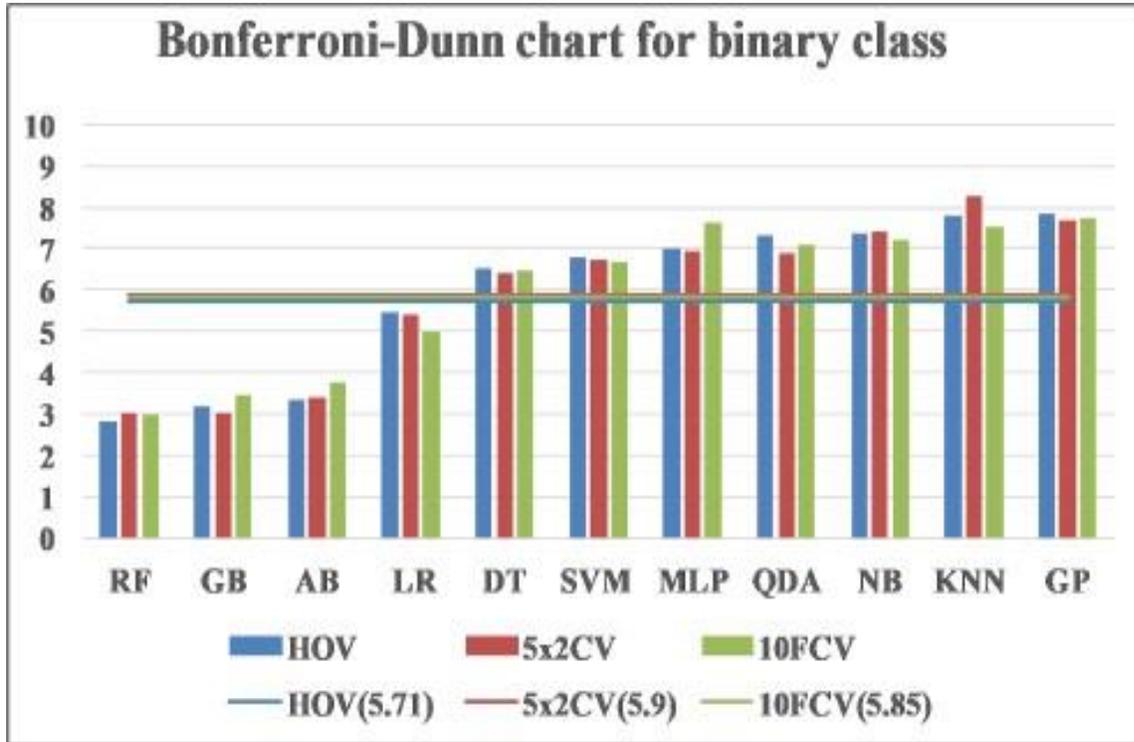


FIGURE 2. BONFERRONI-DUNN CHART FOR BINARY DATA

Figure 2 shows that, when using all cross validation approaches, the performance of DT, SVM, MLP, QDA, NB, and KNN for binary datasets is inferior to that of RF. Figure 3 depicts the bonferroni-dunn graph for multiclass datasets in the same way.

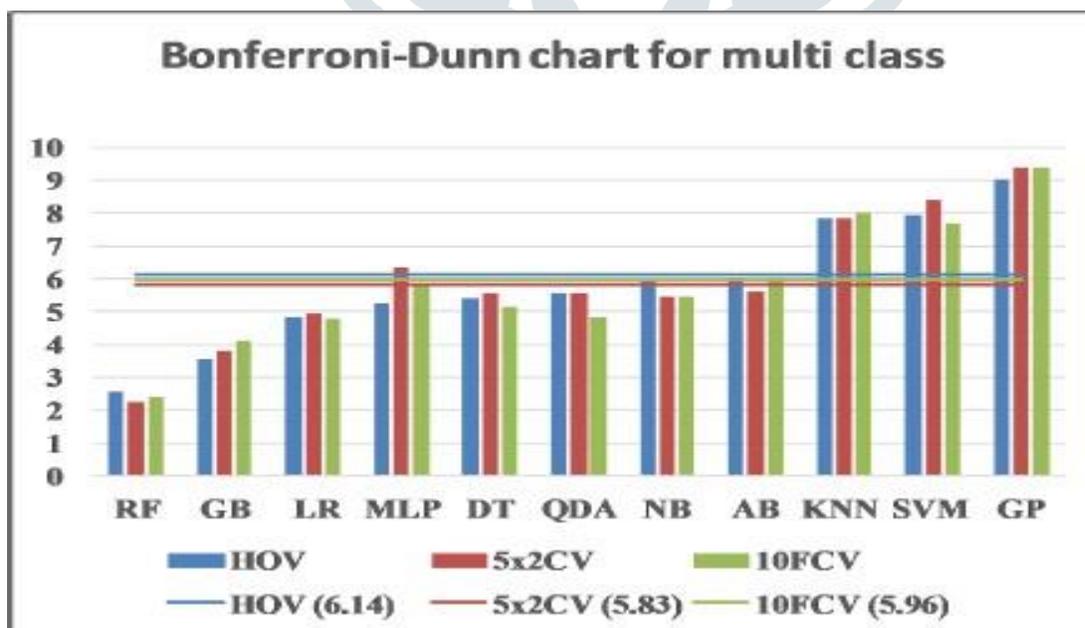


FIGURE 3. BONFERRONI-DUNN CHART FOR MULTICLASS DATA

Figure 3 shows that the algorithms RF, GB, and LR outperformed the others. We have not modified any of the parameters for any algorithms except for K-NN, where the values of K used are equal to 5 because $k=5$ has demonstrated the highest performance. This work aims to discover a generic model that may be extensively utilised in applications utilising internet sensors. For the value of K, the trials are conducted from 1 to 5.

Because it makes a careful final decision based on decisions taken from its component weak base classifiers, (ELs) have given better performance than single learning models for binary class and multiclass. It combines the advantages of several weak models. (RF) operates on both data and feature space at the same time. To accurately categorise the instance, it samples both the training sample space and the feature space. RF is utilised ensemble approach in the health-care arena due to its simplicity and interpretability

1.5. ENSEMBLES OF DECISION TREE FOR ONLINE DATA ANALYTICS

Building a multiclassifier system or having classifiers is a good and efficient technique to handle complicated data. It essentially mixes the hypotheses of many or disparate classifiers in order to generate a more accurate approximation of the genuine hypothesis.

Researchers have been successfully establishing the benefits of (EL) for learning tasks due to the resilience and efficiency of ensemble approaches. Information retrieval, image classification, financial domain, sentiment analysis, natural language processing, online dynamic security assessment, and medical domain have all demo the effectiveness of EL. Not only did EL do well in multiclass classification, but it also did well in multi-label classification. For information extraction, early efforts created learning strategies such as bagging, boosting, and stacking. Nature allows for both homogeneous and diverse ensembles. In big data, ELs are also employed as preprocessing techniques to eliminate noisy instances from data before converting it to smart data. Researchers have also looked studying how feature selection approaches and different base classifiers affect EL.

1.6. E M OF DECISION TREE

1. (BaGing): The bagged decision tree is an ensemble decision tree approach. It produces N bags at random from the original dataset, and N decision trees are trained for these N bags, with the average of these N decision trees utilised as the outcome in the end. An overfitting problem may be avoided by using a constrained decision tree to decrease variance.
2. (AdB): A boosting method that improves on the performance of previously learnt models. As the basis learner, other algorithms may be utilised. It has a high level of accuracy.

3. (GB): It is a boosting technique improvement. GB employs gradient descent to optimise the loss function. A group of trees is created one after the other, and then the individual trees are together in that sequence. Each consecutive trained tree optimizes/reduces the preceding tree's losses.
4. (RF) is a decision tree-based supervised ensemble technique. It employs bagging. The approach takes a random selection of "n" features and trains an individual decision tree on each feature set. RF may be parallelized. It is not affected by liers or non-linear data.
5. ET : It's called the ensemble algorithm. It employs the original dataset rather than subsets. It separates nodes while creating the tree by picking cut sites at random. Variability is reduced by using an extremely (ET). When compared to other bagging and (RF) ensembles, it takes less time to build the ensemble.

1.7. FEATURE REDUCTION TECHNIQUES

1.7.1. PRINCIPLE COMPONENT ANALYSIS (PCA)

This is a linear transformation algorithm that is unsupervised. It generates new components, known as Principal Components, in a low-dimensional subspace (PCs). These PCs are projected orthogonally in order to maximise data variance. The obtained PCs are listed in decreasing order of variance.

1.7.2. LINEAR DISCRIMINANT ANALYSIS (LDA)

It is a supervised and linear feature reduction approach. It searches for new feature space into which to project data in order to increase class separability. LDA converts a high-dimensional feature into a lower-dimensional space while retaining critical information for data categorization.

It condenses the characteristics to a number less than or equal to the number of classes-1.

1.7.3. ISOMETRIC FEATURE MAPPING (ISOMAP)

It calculates the distance between two spots that are far apart without taking surrounding points into account. The measured distance is known as the geodesic distance, and it is used to create a low-dimensional space.

2. RESULTS

2.1. ACCURACY

- Binary: The performance difference between LDA and PCA for all methods ranges from \square 2.5 to 4.5 percent, with the exception of gradient boost and AdaBoost, which are both 0.5 percent. The difference between IsoMap and PCA is 1.5 to 2.0 percent. PCA, Gradient boost, and AdaBoost have comparable performance and are the top scorers. They differ from other algorithms by about 1%.
- Multiclass: The observed difference in performance for algorithms ranging from PCA to LDA is around 3% to 7%, which is substantial. The difference between PCA and IsoMap is between 1% and

2%. Using PCA, ET and RF fared better with comparable performance and are the top scorers. They differ from other algorithms by around 2.55 percent to 7.55 percent, which is much more.□

2.2. PRECISION

- Binary: The reported difference in performance for algorithms ranging from PCA to LDA ranges from 2.5 to 4.5, with the exception of gradient boost and AdaBoost, which are both □ 0.5. The difference between IsoMap and PCA is between 1.5 and 2.5. Using PCA, gradient boost and adaBoost work similarly and get high scores. They differ from the others by around one.
- Multiclass: The observed difference in performance for algorithms ranging from PCA to LDA is considerable and ranges from 3 to 7. Their ratio of PCA to IsoMap is 0.5 to 2. Using PCA, ET and RF outperformed with comparable performance and are high scorers. They differ from others in a range of around 3.0 to 7.0.□

2.3. RECALL

- Binary: The reported difference in performance for algorithms ranging from PCA to LDA ranges from 2.5 to 4.5, with the exception of gradient boost and AdaBoost, which are both □ 0.5. The difference between PCA and IsoMap is between 1.5 and 2.0. Using PCA, gradient boost and adaBoost work similarly and get high scores. They differ from the others by around one.
- Multiclass: The observed difference in algorithm performance between PCA and LDA is considerable and ranges from 3 to 7 roughly. Their ratio of PCA to IsoMap is 0.5 to 2. Using PCA, ET and RF outperformed with comparable performance and are high scorers. They differ from others in a range of around 3.0 to 7.0.□

2.4. F1-SCORE

- Binary: The reported difference in performance for algorithms ranging from PCA to LDA ranges from 2.5 to 4.5, with the exception of gradient boost and adaBoost, which are both □ 0.5. The difference between PCA and IsoMap varies between 1.5 and 2.0. Using PCA, gradient boost and adaBoost work similarly and get high scores. They differ from the others by around one.
- Multiclass: The observed difference in performance for algorithms ranging from PCA to LDA is considerable and ranges from 3 to 7. Their ratio of PCA to IsoMap is 1.0 to 2.

Using PCA, ET and RF outperformed with comparable performance and are high scorers.□ They differ from others by a factor of about 2.0 to 8.0.

3. CONCLUSION

Following experimenting and charting the findings on bonferroni- dunn charts for binary and multiclass data independently, the following final conclusions may be reached: In the case of binary class data, and adaboost using decision tree as the basis classifier performed better on the accuracy scale than other classifiers utilised. Logistic regression has provided decent results but is close to the cutoff value. In the case of multiclass data, random forest and gradient boosting performed better in terms of accuracy. Logistic regression has also performed well, although it is closer to the threshold. It may also be deduced that not all algorithms perform similarly on all types of issues. They outperform other ML algorithms substantially. As a result, EM should be chosen over individual algorithms when no historical data distribution information is available. With these findings, we will investigate the impact of alternative ensembles of (DT) on the IoT data set. It is also vital to investigate how FRS affects other performance measures.

After analysing the findings and comparing performance metric values, it is possible to infer that gradient boost and adaBoost should be favoured over others for binary classification with PCA or LDA. They regularly scored well in terms of accuracy, precision, recall, and F1- score. Extremely randomised tree and Random forest techniques should be selected over other multiclass classification algorithms that employ PCA as a reduction approach. When compared to the others, ET and RF performed better. They regularly scored well in terms of accuracy, precision, recall, and F1-score.

4. REFERENCE

1. R. Baeza-Yates, B. Ribeiro-Neto et al.(1999) Modern information retrieval. ACM press New York, 1999, vol. 463.
2. T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado (2005) “The use of receiver operating characteristic curves in biomedical informatics,” Journal of biomedical informatics, vol. 38, no. 5, pp. 404–415, 2005.
3. A. Ben-David, (2007) “A lot of randomness is hiding in accuracy,” Engineering Applications of Artificial Intelligence, vol. 20, no. 7, pp. 875–885, 2007.
4. J. Huang and C. X. Ling (2005) “Using auc and accuracy in evaluating learning algorithms,” IEEE Transactions on knowledge and Data Engineering, vol. 17, no. 3, pp. 299–310, 2005.
5. H. Yin and N. K. Jha (2017) “A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles,” IEEE Transactions on Multi-Scale Computing Systems, vol. 3, no. 4, pp. 228–241, 2017.
6. A. O. Akmandor and N. K. Jha (2017) “Keep the stress away with soda: Stress detection and alleviation system,” IEEE Transactions on Multi-Scale Computing Systems, vol. 3, no. 4, pp. 269–282, 2017.

7. J. R. Ng, J. S. Wong, V. T. Goh, W. J. Yap, T. T. V. Yap, and H. Ng, (2019) "Identification of road surface conditions using iot sensors and machine learning," in Computational Science and Technology. Springer, 2019, pp. 259–268.
8. N. Dogru and A. Subasi, (2008) "Traffic accident detection using random forest classifier," in 2018 15th learning and technology conference (L&T). IEEE, 2018, pp. 40–45.
9. H. Nguyen, C. Cai, and F. Chen (2017) "Automatic classification of traffic incident's severity using machine learning approaches," IET Intelligent Transport Systems, vol. 11, no. 10, pp. 615–623, 2017.
10. A. Asuncion and D. Newman (2007) "Uci machine learning repository," 2007.
11. R. Kohavi et al. (1995) "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Ijcai, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.
12. H. Habibzadeh, A. Boggio-Dandry, Z. Qin, T. Soyata, B. Kantarci, and H. T. Mouftah (2018) "Soft sensing in smart cities: Handling 3vs using recommender systems, machine intelligence, and data analytics," IEEE Communications Magazine, vol. 56, no. 2, pp. 78–86, 2018.
13. D. J. Sheskin (2020) Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC, 2020.
14. J. Demšar (2006) "Statistical comparisons of classifiers over multiple data sets," The Journal of Machine Learning Research, vol. 7, pp. 1–30, 2006.
15. M. Friedman (1937) "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," Journal of the american statistical association, vol. 32, no. 200, pp. 675–701, 1937.
16. Friedman (1940) "A comparison of alternative tests of significance for the problem of m rankings," The Annals of Mathematical Statistics, vol. 11, no. 1, pp. 86–92, 1940.
17. R. L. Iman and J. M. Davenport (1980) "Approximations of the critical region of the fbietkan statistic," Communications in Statistics-Theory and Methods, vol. 9, no. 6, pp. 571–595, 1980.
18. O. J. Dunn, (1961) "Multiple comparisons among means," Journal of the American statistical association, vol. 56, no. 293, pp. 52–64, 1961.
19. M. Jethanandani, A. Sharma, T. Perumal, and J.-R. Chang (2020) "Multi-label classification based ensemble learning for human activity recognition in smart home," Internet of Things, vol. 12, p. 100324, 2020.
20. D. García-Gil, J. Luengo, S. García, and F. Herrera (2019) "Enabling smart data: noise filtering in big data classification," Information Sciences, vol. 479, pp. 135–152, 2019.
21. M. M. Ghiasi and S. Zendehboudi (2021) "Application of decision tree-based ensemble learning in the classification of breast cancer," Computers in Biology and Medicine, vol. 128, p. 104089, 2021.

22. A. Kim, K. Oh, J.-Y. Jung, and B. Kim (2018) “Imbalanced classification of manufacturing quality conditions using cost-sensitive decision tree ensembles,” *International Journal of Computer Integrated Manufacturing*, vol. 31, no. 8, pp. 701–717, 2018.
23. W. X. Cheng, R. Katuwal, P. N. Suganthan, and X. Qiu (2017) “A heterogeneous ensemble of trees,” in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2017, pp. 1–6.
24. A. Narassiguin, M. Bibimoune, H. Elghazel, and A. Aussem (2016) “An extensive empirical comparison of ensemble learning methods for binary classification,” *Pattern Analysis and Applications*, vol. 19, no. 4, pp. 1093–1128, 2016.
25. W. Hurst, C. A. C. Montanez, N. Shone, and D. Al-Jumeily (2020) “An ensemble detection model using multinomial classification of stochastic gas smart meter data to improve wellbeing monitoring in smart cities,” *IEEE Access*, vol. 8, pp. 7877–7898, 2020.

