QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Reduction in High-Dimensional Data by using HDRF with Random Forest Classifier

Ahmed Najat Ahmed
Department of Computer Engineering, College of Engineering and Computer Science, Lebanese French University, Erbil, Kurdistan Region, Iraq.
a.afandy@lfu.edu.krd

ARTICLE INFO

Article History:
Received: 12/10/2021
Accepted: 25/11/2021
Published: Autumn 2021

Keywords:
Random forest, ensemble pruning, kappa measure, hybrid dimensionality reduction forest.

Doi:
10.25212/lfu.qzj.6.4.30

ABSTRACT

In machine learning, high-dimensional data classification is one of the significant challenges. For dimensionality reduction, the traditional classifiers enhance the variety of classifiers. The conventional method has some restrictions; information loss is caused during dimensionality reduction, minimizing accuracy. The selection of the sample is vulnerable to redundant features and noise. The proposed method, Hybrid Dimensionality Reduction Forest (HDRF) with Random Forest (RF) ensemble classifier and kappa measure, were used to overcome those restrictions. Initially, the Kappa measure is used for pruning, and the higher degree is selected from the forest. For partitioning the features, a tree-based selection method is used. Principal Component Analysis (PCA) is used for feature extraction, noise reduction, and dimensionality reduction. The proposed method removes the weak classifiers and eliminates redundancy. Also, it reduces the unselected structures and the fundamental structures into a new system. The evaluation results on 25 high-dimensional data, the proposed method outperforms with Random Forest ensemble classifier methods and provides enhanced results obtained on 21 out of 25.

1. Introduction

Classification is one of the significant or topics in supervised learning, which is to train a classifier or a collection of classifiers. Different areas such as text classification,

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

medical, image recognition, image and video processing, and ensemble learning are applied. For limited training data, high-dimensional data is very complicated data mining. Ensemble learning reduces the high dimensional data by using feature selection (Lu et al., 2017; Chen et al., 2020).

Traditional ensemble learning algorithms, on the other hand, have the following problems when dealing with high-dimensional data: All of the approaches to classifier ensembles, whether they're focused on increasing the variety of classifiers in a particular classifier space in a specific sample or feature space, are thought to have a positive effect on the diversity of classifiers. The majority of classifier ensemble techniques are utilized for high-dimensional situations. Before dimensionality reduction, the number of features is somewhat increased to increase the variety of classifiers, and most ensemble pruning algorithms have been optimized.

The random forest can deal with continuous and discrete qualities simultaneously and can handle noise better. Furthermore, RF avoids overfitting and is well-suited to dealing with complex situations. Even though it is haphazard, although the forest has numerous benefits, it also has some drawbacks. The value that is missing and the abnormal value. Such is poor categorization with unbalanced data and the inability to regulate a specific model operation (Zhang et al., 2021).

According to the order of the Kappa measure, the base classifier from minor to extensive proposed Kappa pruning. To a considerable extent, the base classifiers with lower Kappa values. The integrated component was chosen. With the assistance of Kappa, after some pruning, we were able to achieve good integration results. Furthermore, the swarm intelligence optimization algorithm can prune and choose selected high-quality decision trees to participate in integration (Margineantu & Dietterich, 1997).

A random forest ensemble classifier is used for removing the weak classifiers and redundant classifiers. Kappa measure is used for pre-pruning along with HDRF. Experiment evaluations have been conducted of 25 high-dimensional data; out of 25, over 21 datasets have outperformed. The contribution of this paper is as follows:

- Unselected samples are used as secondary information to build varied and efficient quality. An ensemble forest pruning approach is employed to examine

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

the impact of the same classifiers on the integrated system based on this information.

- We compare our method to other prominent ensemble learning methods on numerous high-dimensional datasets to assess its effectiveness at deleting redundant classifiers.
- Using the Kappa measure for pre-pruning, the CARTs with poor overall performance are excluded, the range of CARTs is reduced, and the computational burden of included pruning is lowered.

The rest of this research article is written as follows: Section 2 consists of a brief study of HDRF and Random Forest ensemble pruning with kappa measure and describes the working principle of the proposed model. Section 4 evaluates the result and gives the comparison of different algorithms. Section 5 provides the conclusion of the research work.

2. Related works

Using the bootstrap method, every tree is given coaching set with the scale of n. Randomly select M options at nodes compare and choose the most compelling features. Recursively generate each call tree while not pruning (Sheykhmousa et al., 2020). Propose a novel transformative classifier gathering technique called stingy outfit, which accomplishes a trade-off between accuracy and intricacy. Li et al. (2020) propose a recursive troupe learning way to augment the utilization of information in profound learning applications. Dong et al. (2015) offer a classifier blend technique dependent on signal strength, which joins the yield of different classifiers to help dynamic. It uses the assistant data of classifiers in past undertakings to change loads of classifiers (Yan et al., 2020).

It proposes a versatile classifier troupe technique dependent on spatial perception for high-dimensional information, which keeps up with the high execution and variety of classifiers (Wang et al., 2019). Classifier ensemble methods in the second category cognizance on theoretical exploration and analysis of classifier ensemble characteristics. Maximum of associated researches awareness on feature selection

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

and sampling. For instance, He & Cao (2012) study to reduce classifier ensemble via function choice. Wu (2018) presents a characteristic selection technique based on constant-nation multi-goal genetic programming, attaining exquisite performance. By using label relevance as a priori,

Multi-label classification bushes should be examined in MI-woodland to monitor the intrinsic characteristics of label correlation (Diao et al., 2013). Build a dynamic weighted stock selection approach with more than one element based mainly on the XGBoost model (Nag & Pal, 2015). Using Bayesian pruning and Bayesian impartial pruning, a two-step ensemble pruning method is used in layout (Wu et al., 2016). Because PCA may reduce measurement and remove noise highlights, it compresses the unselected weak highlights and the amplified highlights into compact and compensatory features.

The third category primarily applies the classifier ensemble method to diverse application areas. For restorative science, Jiang et al. (2016) propose an outfit learning strategy, which combines a few topsy-turvy stowing outfit classifiers to classify biomedical information. Chen et al. (2009) present a classifier gathering based on versatile blended highlight selection and apply it to epileptic seizure classification. Yu & Ni (2014) use a classifier outfit strategy to decrease the impact of covariates on stride acknowledgment. Alzami et al. (2018) plan skimpy outfit learning for concept discovery in the video. Guan et al. (2014) unravel heterogeneous organize issues by ensemble learning. It analyses the most recent development and existing problems of interruption location technology and proposes a versatile coordinate learning model (Tang et al., 2011). Serafino et al. (2018) To increment the differing qualities of classifiers and propose an outfit strategy based on the optimized sampling strategy, which accomplishes excellent execution.

In the proposed system, the Hybrid Dimensionality Reduction Forest and Random Forest ensemble pruning with kappa measure helps to reduce the high-dimensional data and classifies the irrelevant data. The proposed work is more effective, and it increases the diversity between the classifiers.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

3. Proposed HDRF and Random Forest ensemble classifier with Kappa Measure

High-dimensional data classification is the most critical challenge in machine learning. Without any loss in the information, the reduction in the high-dimensionality data classification is difficult. Therefore, the proposed Hybrid Dimensionality Reduction Forest and Random Forest ensemble classifier with Kappa measure reduces the high-dimensionality data and removes redundant and invalid classifiers. Figure 1 shows the overall architecture of the proposed work.

Figure. 1 shows that the training dataset is used for feature selection, and the pre-pruning kappa measure is used. After selecting the features, it chooses the feature subspace, and then the HDRF-Random Forest ensemble classifier is used. After reducing the noise and unwanted classifiers, then build a tree.

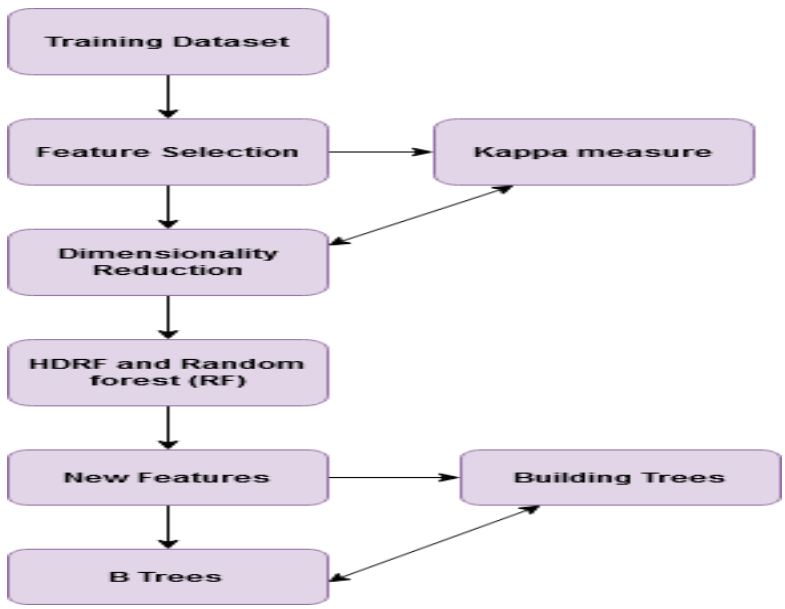

Figure (1): Overall Architecture of HDRF-Random Forest with kappa Measure.

3.1 Hybrid Dimensionality Reduction Forest (HDRF)

HDRF creates a subspace through a powerful and efficient characteristic choice approach called tree-primarily based on distinct selection (Jidong & Ran, 2018). To be

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

more specific, HDRF generates a hundred incredibly random trees using school records before calculating the average significance cost of each characteristic. HDRF also determines the very last ordinary significance of each element by mixing one hundred incredibly random trees using the concept of ensemble learning. Finally, the most critical capabilities are chosen. Tree-based entirely characteristic choice rules pre-selects capabilities with powerful class ability to alleviate the limits of decision tree processing high-dimensional records.

Fore feature selection rate α is used, and N*A is selected for highest importance in the tree-based classifier.

x=[x*Ase , x*(i-Ase)] (1)

Where Ase is denoted as a feature selection matrix, x is a data point.

$$\text{Ase} = \begin{bmatrix} \varphi_1^1 & \varphi_2^1 & \varphi_n^1 \\ \varphi_1^2 & \varphi_2^2 & \varphi_n^2 \\ \varphi_1^n & \varphi_2^n & \varphi_n^n \end{bmatrix} \quad (2)$$

$$\varphi_i^j = \begin{cases} 1 & (i \in \{1, \dots, n\}, j \in \{1, \dots, N\}, \text{xi} \in \text{X}^{\text{s}}) \\ 0 & (i \in \{1, \dots, n\}, j \in \{1, \dots, N\}, \text{xi} \in \text{X}^{\text{u}}) \end{cases} \quad (3)$$

Where X^u is denoted as unselected space and X^s is marked as selected space.

$$\sigma\left(x_i^s, x_j^u\right) = \frac{\sum_{d-1}^{n}(x_{id}^s \times x_{jd}^u)}{\sqrt{\sum_{d-1}^{n}(x_{id}^s)^2} \times \sqrt{\sum_{d-1}^{n}(x_{jd}^u)^2}} \quad (4)$$

Where d ∈ {1,.....,n}, where n is a dimension of the sample, x_{id}^s , and x_{jd}^u is represented as the dth feature of the ith selected sample.

3.2 Kappa measure

The Kappa pruning approach selects the subset of the classifiers. The variety is measured through κ statistic. Set two classifier H1 and classifier H2, facts set containing m examples and l categories, shape a desk wherein cell CEij has x for which H1(x)=i and H2(x)=j.

$$\text{P0}=\frac{\sum_{i-1}^{l} CE_{ij}}{M} \quad (5)$$

$$\text{Pe}=\sum_{i-1}^{l}\left(\sum_{i-1}^{l}\left(\frac{CE_{ij}}{M}\right).\sum_{j-1}^{l}\left(\frac{CE_{ij}}{M}\right)\right) \quad (6)$$

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

P0 is the opportunity that the two classifiers agree (that is, just the sum of the diagonal factors divided through M), Pe is the opportunity that the two classifiers agree with through chance, given the discovered counts withinside the table.

$$K(H1, H2H)=\frac{Po-Pe}{1-Pe} \quad (7)$$

3.3 HDRF and Random Forest with Kappa Measure

Random forest integrates all decision trees to get the final result. It reduces the high-dimensional data and classifies the data correctly. Some low-excellent selection bushes will lessen the accuracy of random woodland. A special RF primarily based on Kappa pruning and HDRF is proposed to enhance the accuracy of random forest. The following steps are random forest steps were shown:

Pseudocode: 1 Random Forest

Step 1: For i=1:n Tree

Step 2: Using the bootstrap method, every tree is given a training set with the dimensions of n. randomly pick out Mtry capabilities at nodes, compare and choose the first-class capabilities. Recursively create every choice tree without pruning.

Step 3: End.

As we all know, the contribution of the various trees within the forest to the algorithmic rule is different, and a few harmful trees could amplify the incorrect prediction, which can cut back the prediction performance of the forest. Therefore, sub-forests obtained by pruning some relatively harmful trees have higher performance than total forests.

Algorithm 1: HDRF and Random Forest with Kappa Measure

Input: x,y,b,α, training set, testing set

Output: optimal tress ad its accuracy

Start

Step 1: To construct a selection matrix, it adopted a tree-based feature selection algorithm in (2).

Step 2: for i=1..... b

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Step 3: For each xsid in Xs (for i =1...n):
Step 4: For each xujd in Xs (for j =1...u):
Step 5: Calculate the similarity σ(xsi,xuj) as in equation 4
Step 6: Pre-pruning is calculated by eq 6
Step7: classifier is built from the training set
Step 8: PCA is used for noise reduction
Step 9: optimal tress are calculated
Step 10: End

Thus, algorithm 1 shows the dimensionality reduction using HDRF and Random Forest with Kappa measure, classifying the optimal solution. PCA reduces the noises and unwanted classifiers.

4. Implementation Results and Analysis

The evaluation result of the proposed system reduces the high-dimensionality data, and it is evaluated by using 25 high-dimensional datasets from the real world. The proposed HDRF and Random Forest with Kappa measure use the following metrics for evaluation: accuracy, precision, recall, and pruning rate.

4.1 Accuracy

It is used to evaluate the classification of the training dataset and classify the dataset correctly. For accuracy calculation, it works with existing classifiers j48, KNN, ET, and the proposed classifier RF outperforms better.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} X100 \quad (8)$$

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

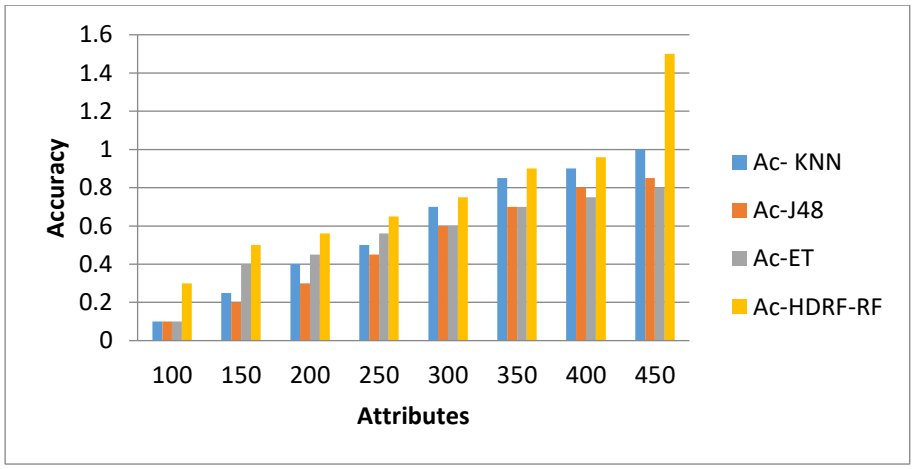

Figure (2): Accuracy comparison

Figure 2 shows that the dataset takes 450 attributes the proposed HDRF and Random Forest achieves 1.75% accuracy.

4.2 Effect of pruning rates

The different pruning rates are evaluated, the x-axis represents the average accuracy rate, and the y-axis represents the pruning rate. It increases the pruning rate from 0.32 to 0.45, and it lost the average accuracy is 0.003 to 0.015.

Figure 3 shows that the proposed random forest outperforms a better pruning rate.

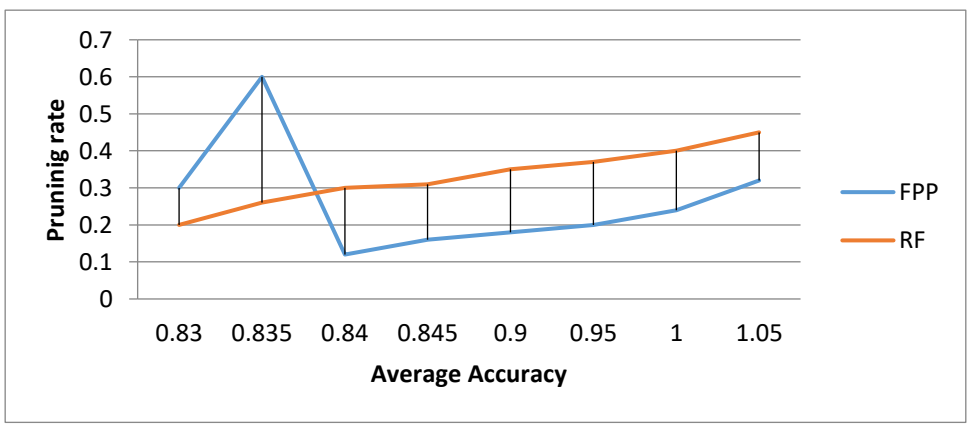

Figure (3): Effect of pruning rates

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

5. Comparison with Previous Works

The proposed classifier is compared with previous well know classifiers such as J48, KNN, and ET. The comparison terms used are precision and recall. Precision means it correctly classifies the dataset and gives the minimized dimensions provided by the following equation:

$$precision = \frac{TP}{TP+FP} X100 \quad (9)$$

Recall means From the total classification data, it selects and classifies the data correctly. The following equation is used for evaluation.

$$recall = \frac{TP}{TP+FN} \quad (10)$$

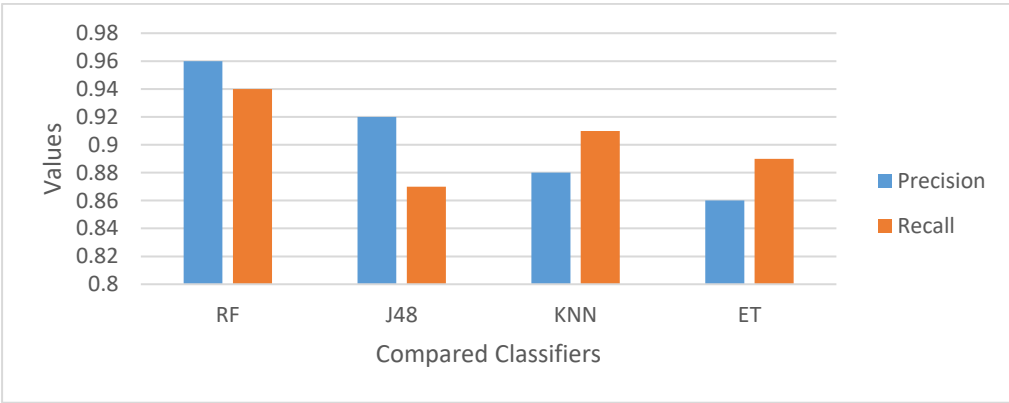

Figure (4): Precision and Recall Values

From the given figure 4, it is evident that the proposed RF classifier outperforms all the other well-known classifiers in the field of study.

6. Conclusion

High dimensional data reduction is one of the significant challenges in Machine learning. In the existing system, it uses FPP for pruning. The proposed method uses HDRF and random forest with Kappa measure to help reduce the high-dimensional data. PCA helps to remove the noises and irrelevant classifiers. Kappa measure helps in the pre-pruning process and improves the pruning rate. Unselected information is taken as auxiliary information and brings new features. The new features are given into HDRF and Random Forest for data classification. Random forest ensemble

classifier removes the redundant classifier of the system. The ensemble method on 25 high-dimensional datasets performs 21 datasets effectively.

In the future, Quadratic Margin Maximization (QMM) will be used for effective pruning and improve the dimension reduction performance. SPCA will be used for noise reduction.

References:

1. Alzami, F., Tang, J., Yu, Z., Wu, S., Chen, C. P., You, J., & Zhang, J. (2018). Adaptive hybrid feature selection-based classifier ensemble for epileptic seizure classification. *IEEE Access*, *6*, 29132-29145.
2. Chen, H., Tiňo, P., & Yao, X. (2009). Predictive ensemble pruning by expectation propagation. *IEEE Transactions on Knowledge and Data Engineering*, *21*(7), 999-1013.
3. Chen, W., Xu, Y., Yu, Z., Cao, W., Chen, C. P., & Han, G. (2020). Hybrid dimensionality reduction forest with pruning for high-dimensional data classification. *IEEE Access*, *8*, 40138-40150.
4. Diao, R., Chao, F., Peng, T., Snooke, N., & Shen, Q. (2013). Feature selection inspired classifier ensemble reduction. *IEEE transactions on cybernetics*, *44*(8), 1259-1268.
5. Dong, Y., Du, B., & Zhang, L. (2015). Target detection based on random forest metric learning. *IEEE Journal of selected topics in applied earth observations and remote sensing*, *8*(4), 1830-1838.
6. Guan, Y., Li, C. T., & Roli, F. (2014). On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence*, *37*(7), 1521-1528.
7. He, H., & Cao, Y. (2012). SSC: A classifier combination method based on signal strength. *IEEE Transactions on neural networks and learning systems*, *23*(7), 1100-1117.
8. Jiang, Z., Liu, H., Fu, B., & Wu, Z. (2016). A novel Bayesian ensemble pruning method. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, IEEE,1205-1212.
9. Jidong, L., & Ran, Z. (2018). Dynamic weighting multi-factor stock selection strategy based on XGboost machine learning algorithm. In *2018 IEEE International Conference of Safety Produce Informatization (IICSPI),* IEEE, 868-872.
10. Li, Y. S., Chi, H., Shao, X. Y., Qi, M. L., & Xu, B. G. (2020). A novel random forest approach for imbalance problem in crime linkage. *Knowledge-Based Systems*, *195*, 105738.

11. Lu, Y. C., Lu, C. J., Chang, C. C., & Lin, Y. W. (2017). A hybrid of data mining and ensemble learning forecasting for recurrent ovarian cancer. In *2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS),* 216-216. IEEE.
12. Margineantu, D. D., & Dietterich, T. G. (1997). Pruning adaptive boosting. ICML'97, 211-218.
13. Nag, K., & Pal, N. R. (2015). A multiobjective genetic programming-based ensemble for simultaneous feature selection and classification. *IEEE transactions on cybernetics*, *46*(2), 499-510.
14. Serafino, F., Pio, G., & Ceci, M. (2018). Ensemble learning for multi-type classification in heterogeneous networks. *IEEE Transactions on Knowledge and Data Engineering*, *30*(12), 2326-2339.
15. Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine vs. random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308-6325.
16. Tang, S., Zheng, Y. T., Wang, Y., & Chua, T. S. (2011). Sparse ensemble learning for concept detection. *IEEE Transactions on Multimedia*, *14*(1), 43-54.
17. Wang, Y., Yang, Y., Liu, Y. X., & Bharath, A. A. (2019). A recursive ensemble learning approach with noisy labels or unlabeled data. *IEEE Access*, *7*, 36459-36470.
18. Wu, O. (2018). Classifier ensemble by exploring supplementary ordering information. *IEEE Transactions on Knowledge and Data Engineering*, *30*(11), 2065-2077.
19. Wu, Q., Tan, M., Song, H., Chen, J., & Ng, M. K. (2016). ML-FOREST: A multi-label tree ensemble method for multi-label classification. *IEEE transactions on knowledge and data engineering*, *28*(10), 2665-2680.
20. Yan, S., Ye, L., Han, S., Han, T., Li, Y., & Alasaarela, E. (2020). Speech Interactive Emotion Recognition System Based on Random Forest. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, IEEE, 1458-1462.
21. Yu, H., & Ni, J. (2014). An improved ensemble learning method for classifying high-dimensional and imbalanced biomedicine data. *IEEE/ACM transactions on computational biology and bioinformatics*, *11*(4), 657-666.
22. Zhang, C., Wang, X., Chen, S., Li, H., Wu, X., & Zhang, X. (2021). A Modified Random Forest Based on Kappa Measure and Binary Artificial Bee Colony Algorithm. *IEEE Access*, *9*, 117679-117690.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

پـوختـه:

له فێربوونی ئامێریدا، پۆلێنکردنی داتا و زانیاریی ڕەهەندبەرز بە یەکێك لە بەرەنگارییە سەرەکییەکان دەژمێردرێت. بەئامانجی کەمکردنەوەی ڕەهەندەکان، پۆلێنکەرە تەقلیدییەکان ئاستی هەمەجۆریی پۆلێنکەرەکان بەرز دەکاتەوە. میتۆدی تەقلیدی ژمارەیەك کۆتوبەندی هەیە، بەشێوەیەك کە دەبنەهۆی لەدەستدانی زانیاری لەکاتی کەمکردنەوەی ئاستی ڕەهەندەکان. هەربۆیەیش لەئاکامدا ئاستی وردی کەم دەکاتەوە. هەڵبژاردنی ئەو نموونەیەی کە ئەگەری تووشبوونی بە خەسڵەتەکانی دووبارەبوونەوە و ژاوەژاوە هەیە. وە لەپێناو بەزاندنی ئەم کۆتوبەندانە، میتۆدی پێشنیارکراوی (دارستانی کەمکردنەوەی ڕەهەندی هایبرید - Hybrid Dimensionality Reduction Forest (HDRF)) بەکارهێنراوە لەگەڵ بەکارهێنانی پۆلێنکەری گرووپی (دارستانی ڕەمەکی - Random Forest (RF)) و پێوانەی کاپا (Kappa). لەسەرەتادا، پێوانەی کاپا (Kappa) بۆ داتاشین بەکاردەهێندرێت و بەرزترین پلە لە دارستانەکە هەڵدەبژێردرێت. بەمەبەستی دابەشکردنی خەسڵەتەکاندا، میتۆدی هەڵبژاردن لەسەر بنەمای درەخت بەکاردەهێندرێت. (شیکاری پێکهاتە سەرەکییەکان - PCA) بەکاردەهێندرێت بۆ دەرهێنانی خەسڵەتەکان و کەمکردنەوەی ئاستی ژاوەژاوە و کەمکردنەوەی ڕەهەندەکان، هەروەها دەبێتەمایەی کەمکردنەوەی ژێرخانە هەڵنەبژێردراوەکان و ژێرخانە ڕەهاکان لە ژێرخانێکی نوێ. میتۆدی پێشنیارکراو هەڵدەستێت بە فرێدانی پۆلێنکەری ناکارا و دووبارەبوونەوەکان فرێ دەدات. ئەنجامەکانی هەڵسەنگاندن لەسەر 25 داتا و زانیاریی ڕەهەندبەرز و میتۆدی پێشنیارکراو کاراترن لە میتۆدەکانی پۆلێنکردنی گرووپی (دارستانی ڕەمەکی) و ئەنجامی کاریگەر بەدەست دێنن لە ئەنجامی 21 لە کۆی 25.

الملخص:

في التعلم الآلي، يعد تصنيف البيانات عالية الأبعاد أحد التحديات الرئيسية. ولتقليل الأبعاد، تعمل المصنفات التقليدية على تعزيز تنوع المصنفات. إن للطريقة التقليدية بعض القيود، حيث يحدث فقدان المعلومات أثناء تقليل الأبعاد. ولذلك، فإنه يقلل من مستوى الدقة. إختيار العينة عرضة للسمات المتكررة والضوضاء. ومن أجل التغلب على هذه القيود، تم استخدام الطريقة المقترحة (غابة تقليل الأبعاد الهجينة - Hybrid Dimensionality Reduction Forest (HDRF)) مع مصنف مجموعة الغابة العشوائية - Random Forest (RF)) ومقياس كابا (Kappa). في البداية، يتم استخدام مقياس (Kappa) للتقليم، ويتم اختيار الدرجة الأعلى من

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (6), No (4), Autumn 2021
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

الغابة. ولتقسيم الميزات، يتم استخدام طريقة الاختيار القائمة على الأشجار. ويستخدم تحليل المكونات الرئيسية (PCA) لاستخراج الميزات، وتقليل الضوضاء وتقليل الأبعاد، كما أنه يقلل من الهياكل غير المنتقاة والهياكل المطلقة في هيكل جديد. إن الطريقة المقترحة تحذف المصنفات غير الصالحة وتحذف التكرار. نتائج التقييم على 25 بيانات عالية الأبعاد والطريقة المقترحة تتفوق على طرق تصنيف مجموعة الغابة العشوائية وتوفر نتائج معززة تم الحصول عليها في 21 من 25.