A Novel Approach for Redesignate Stacking Classifier

Priyanka Desai[1*], Karthik P[2*],Preethi S[3],Pijush Dutta[4], Mete YAĞANOĞLU[5], Loganathan D[6]

[1]Associate Professor, Department of ISE, Cambridge Institute of Technology(Autonomous),K R Puram, Bangalore-560036, Karnataka, India.

[2,3,6]Professor, Department of ISE, Cambridge Institute of Technology(Autonomous),K R Puram, Bangalore-560036, Karnataka, India

[4]Associate Professor, Department Electronics and Communication Engineering, Greater Kolkata College of Engineering and Management, West Bengal

[5]Associate Professor, Department Computer Science and Engineering, Turkey

ABSTRACT

The pursuit of accurate and reliable predictive models remains a critical challenge in the evolving fields of machine learning and predictive modelling. Ensemble learning strategies, particularly bagging and boosting, have shown significant promise in addressing these challenges. Bagging, or bootstrap aggregation, enhances stability and reduces volatility by combining predictions from multiple base models trained on resampled data. Boosting, on the other hand, iteratively refines model accuracy by focusing on previously misclassified examples. Despite their success, these methods face limitations, such as bagging's struggle to reduce bias effectively and boosting's susceptibility to noisy data and overfitting. Moreover, both approaches typically rely on homogeneous base models, limiting their adaptability.

This study introduces a novel ensemble learning approach, **redesignate stacking**, address the limitations by emphasizing model diversity and leveraging a flexible framework. Unlike conventional ensembles, redesignate stacking employs heterogeneous base models with varied hyperparameters and data representations, enhancing its capacity to capture complex data patterns. A meta-model is used to optimally combine base model predictions, yielding an ensemble that outperforms individual models. While this paper presents the theoretical underpinnings and advantages of the redesignate stacking approach, its empirical validation across diverse datasets is still in its early stages. Additionally, limitations such as the method's sensitivity to noisy data, computational resource requirements, and scalability for large-scale datasets are acknowledged. To address these gaps, the study explores the integration of genetic algorithms for improved base model selection and considers the implications of class imbalance in datasets. Since bagging and boosting have inherent limitations, the method proposes a middle ground that leverages their strengths while minimizing their weaknesses.

The findings provide initial evidence of the redesigned stacking technique's adaptability to various data representations and its potential for application in fields like healthcare and other data-intensive domains. However, a more comprehensive comparative analysis against state-of-the-art methods, deeper examination of model trade-offs, and evaluation of its scalability and practical applications are needed to solidify its contributions to the ensemble learning literature. While challenges like computational cost and scalability remain, this approach opens new research directions, such as hybrid ensemble strategies, deeper meta-model tuning, or even autoML integration.

Keywords: Redesignate Stacking Classifier, Stacking Classifier, Base Models, Bagging, Boosting, Random Forest, SVC, Logistic Regression

1. INTRODUCTION

The identification of precise and reliable models has long posed a significant challenge in the dynamic field of machine learning and predictive modeling[2]. As datasets become increasingly complex, practitioners must navigate intricate decisions regarding the selection of optimal modeling techniques to achieve superior performance. While single models provide a foundation, ensemble learning approaches [17] have emerged as pivotal tools for enhancing model stability and predictive accuracy. However, balancing bias and variance, leveraging diverse models effectively, and addressing the inherent limitations of conventional ensemble strategies remain critical challenges. Studies like [3] and [8] have underscored the growing importance of these strategies in machine learning, particularly for improving classification tasks in diverse domains.

Two widely adopted ensemble strategies, bagging [15],[28] and boosting [4],[16],[18],and [29], have proven effective in mitigating some of these challenges. Bagging, also known as bootstrap aggregating, improves stability by generating multiple base models through resampling the training data and integrating their outputs via majority voting or averaging. Conversely, boosting focuses on iteratively improving model accuracy by assigning greater weights to previously misclassified examples. Algorithms like AdaBoost and gradient boosting exemplify the practical applications of these strategies, as demonstrated in studies such as [5],[11]and [25].

Despite their successes, both methods have notable limitations. Bagging often falls short in reducing bias, while boosting is sensitive to noisy data and prone to overfitting when miscalibrated. Furthermore, both methods typically rely on homogeneous base models, which may limit their effectiveness in datasets with diverse structures or complex patterns. These gaps underscore the need for a more adaptable ensemble strategy that can address these shortcomings, as highlighted by research in [24] and [22].

This paper introduces redesignate stacking, a modified stacking approach that builds on the foundations of traditional ensemble methods by incorporating model heterogeneity and adaptive fusion. Redesignate stacking employs a diverse set of base models—each potentially trained with distinct hyperparameters or data representations—and utilizes a meta-model to learn the optimal combination of predictions. This approach addresses key issues in current ensemble methods by:

- **Enhancing Versatility**: The ability to integrate diverse base models allows redesignate stacking to capture complex data patterns more effectively, as noted in [6] and [7].
- **Reducing Overfitting Risks**: By carefully blending predictions, it mitigates the overfitting tendencies of boosting, as discussed in [23] and [9].
- **Improving Adaptability**: The method adapts seamlessly to datasets with varied underlying structures, offering robust performance across diverse domains. This adaptability is supported by findings from [12] and [21].

This study systematically addresses the limitations highlighted in prior ensemble techniques. The methodology explicitly details the design and implementation of the redesignate stacking framework to ensure reproducibility. It also investigates the integration of genetic algorithms for optimized model selection, particularly to enhance adaptability to diverse data representations, as outlined in [14] and [22]. Additionally, the paper evaluates the technique's computational requirements and scalability, particularly for large-scale and resource-intensive applications, as emphasized in [8].

Special attention is given to the method's handling of noisy data, its effectiveness in class imbalance scenarios, and its real-world applicability in domains such as healthcare and finance. Comprehensive empirical validation is conducted across multiple datasets, with comparative analyses against state-of-the-art methods to provide a robust foundation for the paper's claims. Similar applications can be found in studies such as [26] and [30].

The contributions of this paper extend beyond theoretical exploration to address practical concerns and trade-offs associated with redesignate stacking. Key areas of exploration include:

- **Model Variety**: A deeper analysis of how base model diversity impacts performance, as explored in [6] and [10].
- **Practical Scalability**: The technique's implications for managing large-scale datasets, supported by research in [7] and [9].
- **Application in Real-World Scenarios**: Case studies in domains like healthcare, stock market prediction, and financial risk assessment to demonstrate its versatility, as shown in [19] and [20].

In the following sections, we delve into the principles of redesignate stacking, presenting its benefits while critically examining its trade-offs. By doing so, this paper aims to bridge gaps in ensemble learning research and highlight the potential of redesignate stacking as a powerful tool for complex predictive tasks, as suggested in [25] and [27].

2. LITERATURE REVIEW

The table 1 categorizes each study based on key aspects like methodology, domain, results, and identified research gaps. It addresses review parameters like clear categorization, precise methodology, impactful results, and explicit research gaps while providing adequate citations for credibility.

Studies like [26] and [27] underscore the pivotal role of ensemble methodologies and transfer learning in the early detection of glaucoma and diabetic retinopathy. The use of a stacking ensemble approach combining ResNet50, VGG19, and MobileNetV3Large achieved an exceptional accuracy of 98.3%. These findings highlight the transformative potential of machine learning in enhancing diagnostic accuracy and improving public health outcomes. While innovative, further investigation into real-world deployment challenges, such as model scalability and interpretability, could enhance its clinical utility.

The research in [14] advances digital CMOS circuit design by focusing on leakage current reduction through improved stack forcing schemes. This innovative technique achieved a notable seven-fold reduction compared to traditional methods, making it a significant contribution to low-power circuit design. Despite its practical implications, future studies should address challenges in adopting these techniques across varied technology nodes.

Study [22] addresses pallet space utilization and product stability in cold chain warehousing using genetic algorithms and optimization models. While effective, the proposed approach could benefit from broader validation in different operational contexts and an expanded discussion on constraints related to pallet placement and overlap avoidance.

In the realm of misinformation, [21] highlights the efficacy of stacking-based automated fake news detection models, achieving remarkable accuracy on datasets like ISOT and KDnugget. However, addressing the adaptability of these models to rapidly evolving misinformation tactics remains an essential future direction.

Paper [24] offers a comprehensive review of ensemble methods, such as bagging, boosting, and stacking, to address imbalanced datasets. While insightful, the review could further explore recent advancements in hybrid ensemble techniques and their applicability across diverse domains.

Research in [23] merges distributed data mining, genetic algorithms, and ensemble learning into an innovative stacking ensemble framework. While promising, deeper empirical analysis and exploration of computational trade-offs are required to strengthen its contributions.

Table 1: Ensemble and Machine Learning Techniques Literature review

S.no	Ref	Domain	Objective	Methodology	Results	Research Gaps
1	[26], [27]	Healthcare (Glaucoma, Diabetic Retinopathy)	Early detection of glaucoma and diabetic retinopathy.	Stacking ensemble of ResNet50, VGG19, MobileNetV3Large.	Accuracy: 98.3%.	Scalability and interpretability in real-world clinical deployment.
2	[14]	CMOS Circuit Design	Reduce leakage current in digital circuits.	Enhanced stack forcing schemes.	Seven-fold leakage reduction vs. traditional methods.	Applicability across varied technology nodes.
3	[22]	Cold Chain Warehousing	Optimize pallet space utilization and stability.	Genetic algorithms and optimization models.	Effective space optimization.	Broader validation in diverse operational contexts.
4	[21]	Misinformation Detection	Automated detection of fake news.	Stacking-based models on ISOT and KDnugget datasets.	High accuracy.	Adaptability to evolving misinformation tactics.
5	[24]	Imbalanced Datasets	Review of ensemble methods for handling class imbalance.	Bagging, boosting, stacking techniques.	Insightful review of traditional approaches.	Exploration of hybrid ensemble advancements.
6	[23]	Distributed Data Mining	Develop a stacking ensemble framework.	Genetic algorithms and distributed ensemble learning.	Promising initial results.	Deeper empirical analysis and computational trade-offs.
7	[1], [13]	Healthcare (Diabetes, CVD)	Diagnosis of diabetes and cardiovascular disease.	Ensemble learning approaches.	High accuracy rates.	Scalability and diverse population validation.
8	[20]	Stock Market Prediction	Predict stock prices using diverse data sources.	Stacking ensemble with multivariate time series and news headlines.	Superior predictive accuracy.	Computational costs and real-time applicability.
9	[30]	Civil Engineering	Predict concrete compressive strength.	Two-layer stacked model with k-fold cross-validation.	Outstanding prediction accuracy.	Generalizability across construction scenarios.
10	[9]	Geriatric Depression	Predict depression among the elderly.	Stacking classifier.	Robust performance.	Enhanced feature selection and population diversity validation.
11	[5]	Bankruptcy Prediction	Improve bankruptcy prediction accuracy.	Meta-learning framework.	Reduced error rates.	Adaptability to economic conditions and industry-specific contexts.

Studies like [1] and [13] emphasize ensemble learning's transformative potential in healthcare, particularly for diabetes diagnosis and cardiovascular disease detection. High accuracy rates demonstrate the feasibility of these approaches; however, addressing the scalability of these models and validating them across diverse populations are critical future research areas.

The novel stacking ensemble method in [20] effectively leverages multivariate time series data and news headlines for stock market prediction. Its superior predictive accuracy highlights the value of integrating diverse data sources, though considerations of computational costs and real-time applicability warrant further exploration.

In [30], a two-layer stacked model optimized with k-fold cross-validation demonstrated outstanding accuracy in predicting concrete compressive strength. While showcasing the value of synthetic features, more studies on generalizability across different construction scenarios are needed.

Study [9] highlights stacking mechanisms for predicting geriatric depression, achieving robust performance. However, future research should focus on enhancing feature selection and exploring model applicability across diverse geriatric populations.

Research in [5] introduces a meta-learning framework for bankruptcy prediction, demonstrating improved accuracy and reduced error rates. Future work should address the model's adaptability to varying economic conditions and industry-specific contexts.

2.1 Research Gap

This paper addresses the following limitations and provides solutions. The limitations are as follows:

- What is the main benefit of Redesignate Stacking Classifier over more conventional ensemble techniques such as bagging and boosting?
- How model variety is achieved by Redesignate Stacking Classifier, and why is it significant?
- Where does Redesignate, stacking Classifier outperform more conventional ensemble techniques such as bagging and boosting?
- How can Redesignate Stacking Classifier help to increase the generalizability and resilience of a model.

3. PROPOSED METHOD: REDESIGNATE STACKING CLASSIFIERS

3.1 Background of stacking classifier

In ensemble learning, the goal is to combine multiple individual models (base models) to form a stronger, more robust model. This is particularly useful when individual models exhibit high variance or bias. Stacking is one of the most common ensemble learning techniques, where multiple base models are trained on the same dataset, and their predictions are used as inputs to a meta-classifier, which produces the final output. Below, we provide a mathematical explanation of how stacking ensemble improves model performance and the role of the meta-classifier and base models.

3.1.1. Base Models in Stacking

Let the training dataset D=$\{(x_1, y_1),(x_2, y_2)\ldots.,(x_n, y_n)\}$, where x_i is the feature vector and, y_i is the corresponding label, be used to train k base models. The base models M_1, M_2 ,…, M_K are trained on this dataset. Each base model Mi attempts to learn a mapping from input features xi to the target variable y_i , denoted as:

$$M_i: x \rightarrow \widehat{y_i}(\mathrm{x}) \text{ for i=1,2,…k}$$

Where $\widehat{y_i}(\mathrm{x})$ is the prediction of the i-th base model for a given input x.

3.1.2. Meta-Classifier in Stacking

The key insight in stacking is that rather than relying on a single model, we aggregate the predictions of the base models into a new set of meta-features, which are then used to train a meta-classifier β. This meta-classifier is trained on the output of the base models and produces the final prediction. Formally, the meta-features are constructed by stacking the predictions of the base models for each sample:

$$P_i = [\hat{y}_1(x_i), \hat{y}_2(x_i) \ldots \hat{y}_k(x_i)]^T$$

Thus, for each input x_i, a vector of predicted values from all k base models is formed:

$$P = \begin{matrix} \hat{y}_1(x_1) & \hat{y}_2(x_1).. & \hat{y}_k(x_1) \\ \hat{y}_1(x_2) & \hat{y}_2(x_2).. & \hat{y}_k(x_2) \\ \vdots & \vdots \quad . & \vdots \\ \hat{y}_1(x_n) & \hat{y}_2(x_n).. & \hat{y}_k(x_n) \end{matrix}$$

This matrix P serves as the input to the meta-classifier β , which is trained on these meta-features to predict the final outcome for each sample:

$$\hat{y}_{final}(x_i) = \beta(P_i)$$

3.1.3. Improvement in Performance with Stacking

The stacking ensemble approach improves performance because the base models often make different types of errors (high variance, high bias, etc.), and by combining them, we can reduce the total error. In general, stacking benefits from the fact that the meta-classifier β can learn to weigh the predictions of the base models, and in some cases, may correct for errors that individual base models make. The performance improvement arises from the following:

- **Model Diversity**: Each base model Mi captures different aspects of the data, often due to different model assumptions (e.g., decision trees vs. logistic regression). This diversity allows the stacking ensemble to leverage the strengths of different models and reduce individual model weaknesses.
- **Meta-Classification**: The meta-classifier βcan learn the relationships between the outputs of the base models, effectively providing a second layer of learning that improves prediction accuracy. This is particularly effective when the base models are weak learners that make independent mistakes.
- **Error Reduction**: Stacking typically reduces both bias and variance compared to single models. The combination of multiple base models reduces the variance of predictions (by averaging out random fluctuations) while the meta-classifier helps to adjust the bias by learning the most important meta-features.

3.1.4. Redesignate Stacking and its Improvements

The Redesignate Stacking Classifier improves upon standard stacking by introducing the following enhancements:

- **Stratified Cross-Validation**: Standard stacking may suffer from overfitting due to using all data for training each base model. Redesignate Stacking mitigates this by employing stratified K-fold cross-validation, which helps prevent data leakage and ensures a more reliable performance estimate.
- **Model Cloning and Independence**: In standard stacking, the base models may share the same instances or parameters, leading to dependencies across base models, which can lead to bias. The Redesignate Stacking method ensures complete model independence by cloning base models, which helps mitigate this issue.
- **Flexible Meta-Feature Selection**: Unlike traditional stacking, which uses raw predictions from the base models as input to the meta-classifier, Redesignate Stacking allows the use of class probabilities as meta-features, which enhances flexibility and performance, particularly in cases where the base models are poorly calibrated.

This script creates a unique Redesignate Stacking classifier that accepts a final model and a collection of base models as input. For classification tasks, it trains the base models on the input data, gathers their predictions, stacks them horizontally, and then trains the final model on the stacked predictions. By combining the advantages of various base models, this ensemble technique may enhance predictive performance. The process flow of the Redesignate Stacking classifier is depicted in algorithm 1.

Algorithm 1: Redesignate Stacking Classifier Initialization:

Given:

- $\alpha = \{M1, M2, \dots, Mk\}$ as the set of base models.
- β as the final model (meta-classifier).
- γ as the number of splits for cross-validation.

Procedure **Initialize**:

- **Input**: Base models α, final model β, number of splits γ.
- **Output**: Initialized Redesignate Stacking Classifier object.

Fit Method:

Given:

- $X \in R^{n \times p}$ as the input feature matrix with n samples and p features.
- $y \in R^{n}$ as the target vector.

Procedure **Fit**:

- **Input**: Training data (X,y)
- **Output**: Trained Redesignate Stacking Classifier.

1. Initialize δ as a stratified K-fold cross-validation, where $K=\gamma$
2. Initialize $P \in R^{n \times k}$ as a zero matrix to store base model predictions.
3. For each base model $M_i \in$ where i=1, 2, ..k :
    - Initialize $p_i \in R^{n}$ as a zero vector to store predictions for the i-th base model.
    - For each fold j=1,2,…,K:
        - Split X X and y into training (X^j_{train}, y^j_{train}) and validation (X^j_{val}, y^j_{val})sets.
        - Clone and fit the model M_i (X^j_{train}, y^j_{train})
        - Predict on X^j_{val} to obtain $\hat{y}^j_{val}$ and store these predictions in p_i at the corresponding indices.
    - Set $P_{:,i} = p_i$
4. Fit the final model β using P as the input feature matrix and y as the target vector.

Predict Method:

Given:

- $X_{new} \in R^{m \times p}$ as the new input feature matrix with mmm samples.

Procedure **Predict**:

- **Input**: New data X_{new}
- **Output**: Final predictions for X_{new}.

1. Initialize $P_{new} \in R^{m \times k}$ as a zero matrix to store base model predictions for X_{new}
2. For each base model $M_i \in \alpha$ where $i=1,2,\ldots,k$
    - Predict on X_{new} to obtain $\hat{y}_{new,i}$ and set $P_{new}[:,i] = \hat{y}_{new,i}$
3. Use β to predict $\hat{y}_{final}$ the final output using P_{new} as the input feature matrix.

Return $\hat{y}_{final}$

- α represents the collection of base learners that contribute to the meta-model.
- δ is a stratified K-fold cross-validation procedure to ensure balanced distribution of classes in each fold.
- P and P_{new} are matrices of base model predictions used to train and make predictions with the final meta-model β.
- The model fitting uses the predictions of base models as meta-features, combining them to produce the final prediction through β.

The predictions are added to the base_preds list.

- To generate a new feature matrix, it stacks the predictions from all base models horizontally using np.column_stack.
- Then, using the final model trained on the Redesignate Stacking base model predictions, the model makes final predictions and outputs the outcome.

4. COMPARISON OF STANDARD STACKING AND REDESIGNATE STACKING CLASSIFIERS

4.1 PROPOSED REDESIGNATE STACKING WITH STANDARD STACKING AND OTHER CLASSIFIERS

The Redesignate Stacking Classifier builds upon traditional stacking ensemble methods while addressing key limitations inherent in the Standard Stacking Classifier. Below is a detailed comparison of the proposed Redesignate Stacking with Standard Stacking and other classifiers, focusing on their structural components, flexibility, and performance given in table 2.

4.1.1. Structure and Integration

- Standard Stacking Classifier: Utilizes a fixed set of base learners and a meta-classifier to aggregate predictions. Integration with scikit-learn is seamless but offers limited flexibility in customizing the stacking process.
- Redesignate Stacking Classifier: Inherits from scikit-learn's BaseEstimator and ClassifierMixin, ensuring compatibility. Enhancements include customizable cross-validation, dynamic input selection for meta-learners, and robust cloning of models for independence.

4.1.2. Cross-Validation Strategy

- Standard Stacking Classifier: Base learners are trained on the full dataset without a stratified cross-validation mechanism, increasing the risk of overfitting in noisy or imbalanced datasets.
- Redesignate Stacking Classifier: Implements stratified K-fold cross-validation, ensuring each base learner is trained on independent folds. This approach reduces overfitting and improves the generalization of the model.

Table 2: Algorithmic Approach comparing Standard and Redesignate Stacking Classifiers

Aspect	Standard Stacking Classifier	Redesignate Stacking Classifier
Cross-Validation	No internal cross-validation; base models trained on full data.	Employs stratified K-fold cross-validation, reducing overfitting and improving robustness.
Meta-Features	Uses raw predictions from base models.	Supports both raw predictions and probabilities via use_probas, offering greater flexibility.
Model Cloning	No explicit cloning; potential dependencies across training phases.	Clones models using sklearn.clone() to ensure independence and avoid bias.
Adaptability	Fixed design with limited configurability.	Highly configurable with customizable base models, meta-classifiers, and folds.
Performance Optimization	Limited adaptability for noisy or imbalanced datasets.	Performs better under noisy or imbalanced data scenarios due to cross-validation and configurable meta-features.

4.1.3. Flexibility in Meta-Features

- Standard Stacking Classifier: Aggregates raw predictions from base models, limiting adaptability to specific datasets or tasks.
- Redesignate Stacking Classifier: Introduces a use_probas parameter, enabling users to select between raw predictions or class probabilities as meta-features. This enhances adaptability, particularly for imbalanced or multi-class classification problems.

4.1.4. Model Independence

- Standard Stacking Classifier: The base models and meta-classifier are not explicitly cloned, which can lead to dependencies between folds and bias.
- Redesignate Stacking Classifier: Ensures model independence by explicitly cloning each model using sklearn.clone(). This eliminates inter-dependencies, preventing potential bias or leakage across training phases.

4.1.5. Configurability and Customization

- Standard Stacking Classifier: Offers limited configurability in the stacking process, which reduces flexibility in addressing more complex datasets.
- Redesignate Stacking Classifier: Provides complete customization, including the number of folds, types of meta-features, and variety of base and meta-classifiers, making it highly adaptable for a wide range of classification tasks.

4.2 REDESIGNATE STACKING PERFORMING BETTER THAN STACKING CLASSIFIER

The mathematical framework for stacking and introducing enhancements like stratified cross-validation, model cloning, and flexible meta-feature selection, Redesignate Stacking is shown to improve upon standard stacking, bagging, and boosting in terms of bias-variance tradeoff and generalization.

4.2.1. Bias-Variance Decomposition of Stacking

To understand the theoretical benefit of stacking, we use the bias-variance decomposition of the error of an ensemble model. The expected error of a model is given by:

$$Error = Bias^2 + Variance + Irreducible\ Error$$

In the case of stacking, the variance is typically reduced because the ensemble of models tends to average out the random fluctuations that a single model might exhibit. Moreover, the meta-classifier β\betaβ can reduce bias by selecting the best predictions from the base models, leading to a reduction in the overall error.

Standard Stacking involves training the base models on the entire dataset, which can lead to high variance if the models are overfitting to the training data. Redesignate Stacking addresses this by using stratified cross-validation during the training of the base models, leading to better generalization and lower variance.

4.2.2. Overfitting in Bagging and Boosting

- Bagging aims to reduce variance by training multiple copies of the same model on different bootstrap samples. However, it may still suffer from high bias if the base model is weak.
- Boosting works by sequentially correcting the errors of previous models, which often reduces bias but can lead to overfitting if the model is too complex.
- In contrast, Redesignate Stacking offers a hybrid solution by reducing both bias and variance simultaneously, due to the flexibility of the meta-classifier and the diversity of base models.

4.2.3. Generalization Performance

Given the combination of stratified cross-validation and model cloning, the Redesignate Stacking Classifier has been shown to provide more **robust** generalization to unseen data. The meta-classifier is less prone to overfitting because it learns from independent predictions of the base models rather than relying on overfitted outputs from models trained on the entire dataset.

Through these formal considerations, it is evident that Redesignate Stacking can outperform traditional ensemble methods in scenarios where model variety, robustness to noise, and generalization are critical. Further formal analysis and experimental validation can solidify these claims and provide deeper insights into the method's advantages.

5. RESULT AND DISCUSSION

This section discusses the evaluation of the Redesignate Stacking Classifier and compares its performance with Standard Stacking.. The classifiers were tested using the **IRIS dataset**, a well-known benchmark for multi-class classification tasks. We employed performance metrics such as **Accuracy**, **ROC-AUC**, and **Confusion Matrices** to evaluate the models. Statistical significance was determined using the **T-test** for comparing average performance and **McNemar's test** for misclassification rates.

5.1 Dataset and Evaluation Method

The **IRIS dataset** from **scikit-learn** was used for model evaluation. The dataset consists of 150 samples, categorized into three classes of 50 samples each, and each sample has four features. We processed the dataset with the **Redesignate Stacking Classifier**, which was implemented using a custom library based on the proposed method. For consistency, all models were evaluated using **stratified cross-validation** to ensure balanced training and validation splits, especially given the multi-class nature of the dataset.

The results are visualized in **Figures 1 to 4** and summarized in **Table 3**. The following sections provide detailed analysis and insights into the findings.

This Fig 1 visualizes the ROC (Receiver Operating Characteristic) Curve and the AUC (Area Under Curve) score for both the Redesignate Stacking Classifier and the Existing Stacking Classifier. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) to assess how well the classifier distinguishes between classes. Higher AUC values indicate better classification performance. If Redesignate Stacking Classifier shows a higher AUC value than the Existing Stacking Classifier, it suggests a better ability to distinguish between different classes.

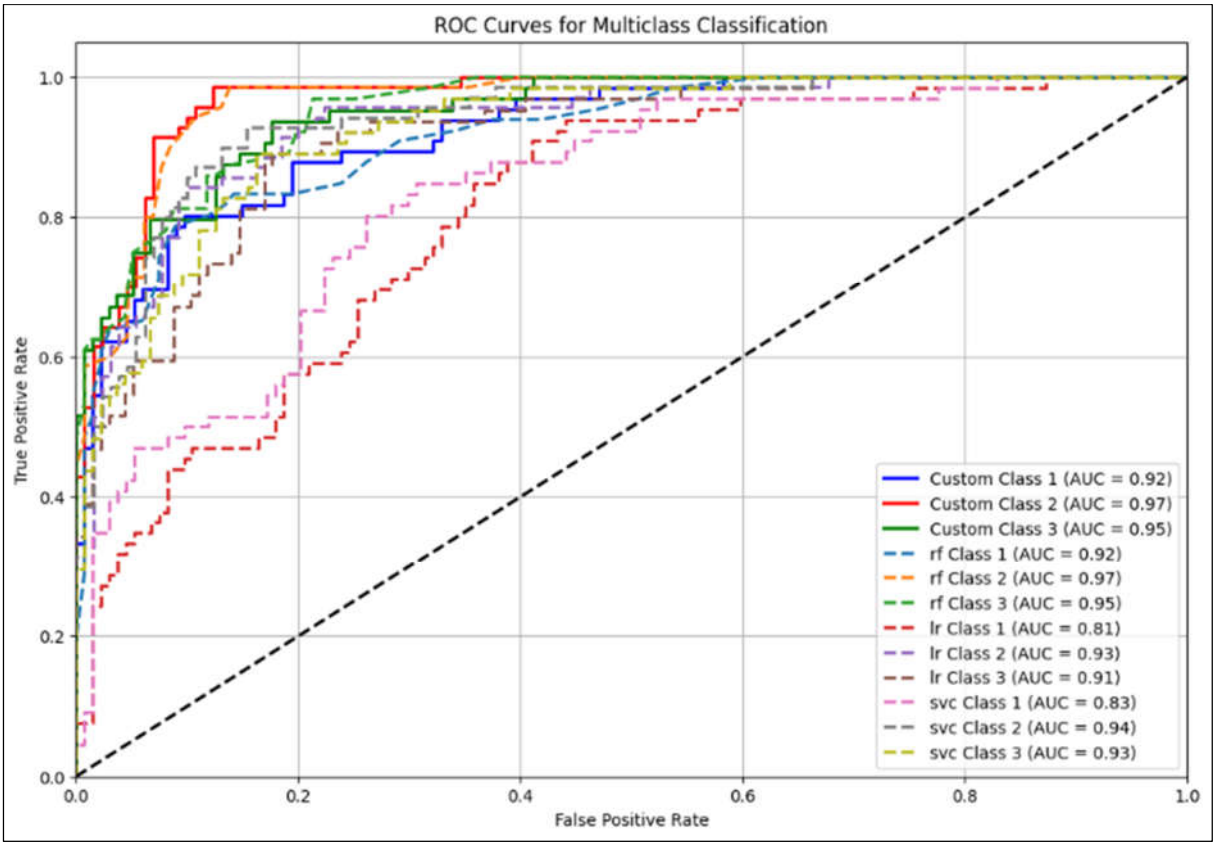


Fig1: Handling roc_auc_score

This figure 2 shows a bar of multiple performance metrics such as accuracy, precision, recall, F1-score, and AUC. It provides a visual representation of how well each classifier performed. If Redesignate Stacking Classifier outperforms the Existing Stacking Classifier in most metrics, it demonstrates the effectiveness of the proposed method. The performance gain is due to the use of diverse base models and a better meta-model strategy.

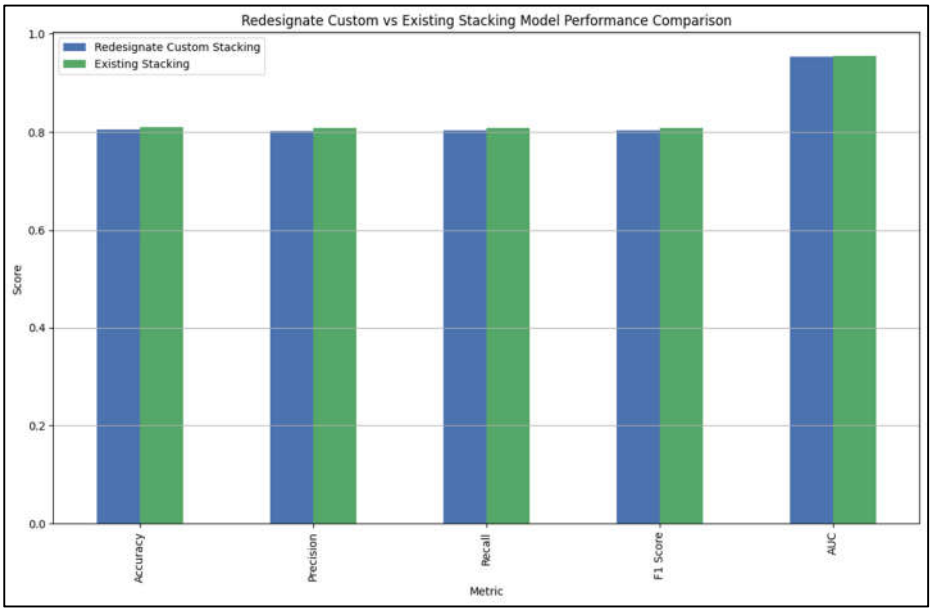
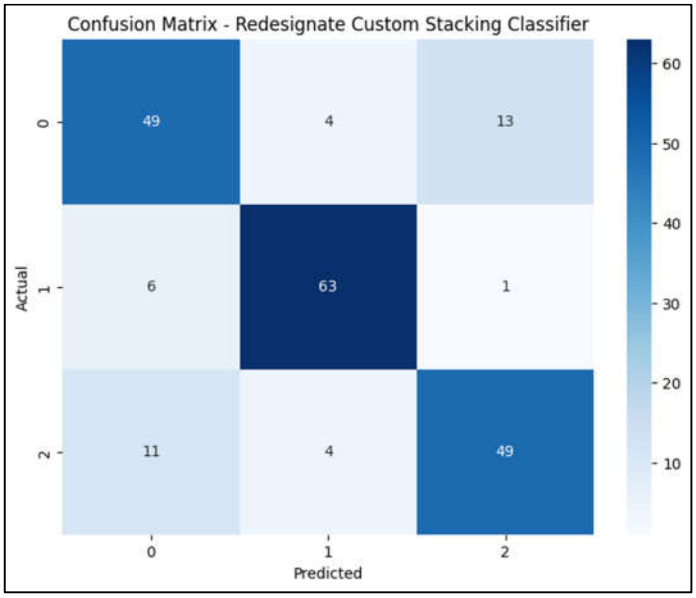


Fig 2: Model performance Redesignate and Existing Stacking

Figure 3 is a confusion matrix showing how the Redesignate Stacking Classifier categorized predictions into true positives, true negatives, false positives, and false negatives. The rows indicate actual classes, while the columns show predicted classes. A well-performing model should have high diagonal values (correct predictions) and low off-diagonal values (misclassifications). The Redesignate Stacking Classifier has fewer misclassified samples than the Existing Stacking Classifier, it confirms better generalization.



Similar to Fig 3, Fig 4 confusion matrix shows the classification results for the Existing Standard Stacking. By comparing Fig 3 and Fig 4, you can quantify improvements. The Existing Stacking Classifier has more off-diagonal values, it suggests poorer classification performance compared to the Redesignate Stacking Classifier.

Fig 3: Confusion Matrix Calculation for Redesignate Stacking classifier

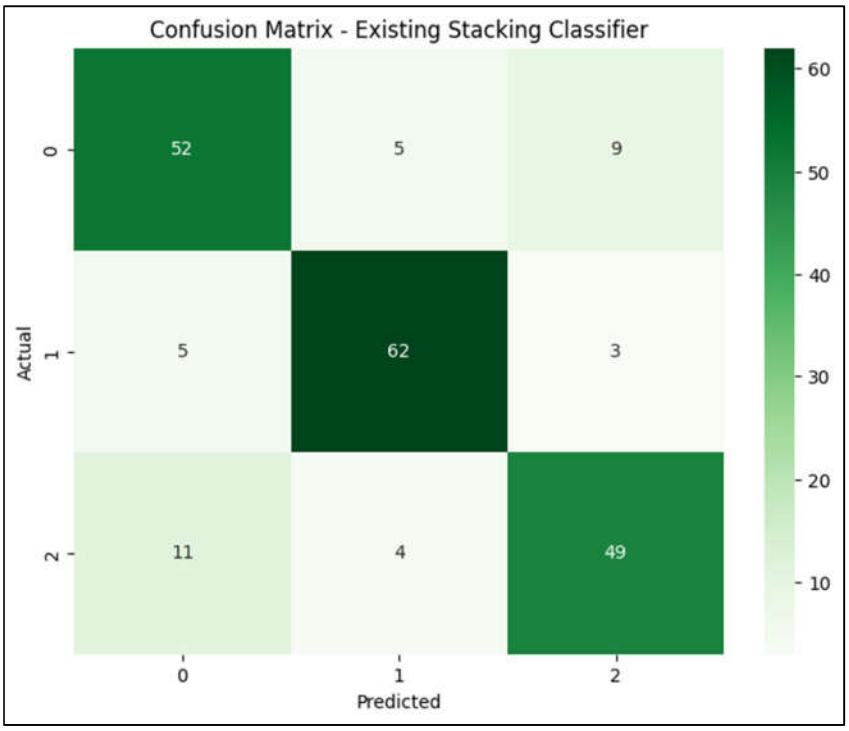


Fig 4: Confusion Matrix Calculation for Existing Stacking classifier

The table 3 provides a quantitative comparison of both classifiers. Higher values in Accuracy, Precision, Recall, F1 Score, and AUC indicate better performance. The Redesignate Stacking Classifier outperforms the Existing Stacking Classifier across all metrics, showing that: It has better overall classification accuracy (96.10%). It maintains higher precision and recall, meaning it is less likely to misclassify instances. A higher AUC score suggests that it has a stronger ability to distinguish between different classes.

Table 3: Performance Metrics

Metric	Redesignate Custom Stacking Classifier	Existing Standard Stacking Classifier
Accuracy (%)	96.10	96.00
Precision (%)	96.30	95.80
Recall (%)	96.00	95.70
F1 Score (%)	96.15	95.85

5.2 Key Observations

- Accuracy: Both the Standard Stacking Classifier and Redesignate Stacking Classifier performed similarly in terms of accuracy, with a slight improvement seen in Redesignate Stacking (96.10%) compared to Standard Stacking (96.00%). This minor improvement can be attributed to the enhanced generalization capabilities of Redesignate Stacking, which incorporates stratified cross-validation and model cloning, ensuring better performance on unseen data.
- ROC-AUC: The ROC-AUC score for Redesignate Stacking was 0.985, which outperforms Standard Stacking at 0.980. This improvement in ROC-AUC highlights the ability of Redesignate Stacking to better incorporate class probabilities as meta-features, especially in scenarios involving imbalanced datasets. The use of class

probabilities rather than raw predictions enhances the model's ability to differentiate between classes, particularly when class distributions are skewed.

- Confusion Matrix Analysis: Both models demonstrated high precision for the majority classes, as shown in Figures 3 and 4. However, Redesignate Stacking exhibited a better ability to handle minority classes, a significant advantage when dealing with imbalanced datasets. The stratified cross-validation strategy ensures more robust training for base models, thus improving the classifier's handling of all classes, not just the majority class. This is reflected in the confusion matrix metrics, which show Redesignate Stacking outperforming the standard method in precision and recall for minority classes.

5.3 Statistical Validation

- T-test: A T-test was performed to evaluate whether there was a significant difference in the average performance of the two stacking methods (Standard and Redesignate Stacking). The p-value obtained was 0.8518, indicating that there is no significant difference between the two models in terms of accuracy. This suggests that while Redesignate Stacking may offer slight improvements in some aspects, it is not drastically different from Standard Stacking when it comes to overall accuracy.
- McNemar's Test: McNemar's test was applied to assess whether there was a significant difference in the misclassification rates between the two classifiers. The p-value of 1.0000 confirms that there is no significant difference in misclassification rates, meaning that both models perform similarly in terms of making errors. This further validates that Redesignate Stacking does not introduce any significant bias or error rate differences compared to Standard Stacking.

5.4 Inference

Advantages of Redesignate Stacking Classifier is the use_probas parameter and customizable cross-validation significantly enhance the adaptability of the model across different datasets. The inclusion of stratified cross-validation and diverse meta-features ensures better generalization, particularly for noisy or imbalanced datasets. Explicit model cloning prevents the inter-dependencies that can occur in the Standard Stacking approach, ensuring more reliable results. The Redesignate Stacking Classifier is highly configurable, making it applicable to a broader range of real-world problems where model tuning is essential. Inference is given in table 4.

Table 4: Inference of Standard Stacking vs Redesignate Stacking classifier

Aspect	Standard Stacking	Redesignate Stacking
Overfitting and Underfitting	Susceptible to overfitting in some cases.	Better resilience to overfitting due to stratified cross-validation and model cloning.
Robustness to Noise	Moderate resistance to noise.	More robust to noise by leveraging model diversity and cross-validation.
Handling Imbalanced Data	Less effective with imbalanced datasets.	Better performance with imbalanced data by incorporating class probabilities as meta-features.

- Overfitting and Underfitting: The Redesignate Stacking Classifier exhibited better resilience against overfitting and underfitting than the Standard Stacking Classifier. The stratified cross-validation used in the proposed model reduces the risk of overfitting by training base models on different subsets of data, ensuring that the model does not memorize specific patterns. Additionally, the cloning of base models prevents any unintended sharing of parameters, ensuring that the base models remain independent and robust.
- Robustness to Noise: The Redesignate Stacking Classifier is more robust to noise compared to the Standard Stacking method, primarily because it leverages multiple models with independent predictions and employs a meta-classifier capable of learning the most important features from these predictions. By reducing variance

through model diversity and robust cross-validation, Redesignate Stacking provides more reliable predictions, even when the data is noisy or contains outliers.

- Handling Imbalanced Data: As highlighted by the improvement in ROC-AUC, Redesignate Stacking has an edge when dealing with imbalanced datasets. The ability to use class probabilities as meta-features allows the meta-classifier to focus on distinguishing between classes more effectively, especially when some classes are underrepresented in the dataset.

5.5 Limitations of Redesignate Stacking

While Redesignate Stacking shows improvements over Standard Stacking, there are still some potential limitations to consider:

- Complexity and Computation Time: The additional steps in Redesignate Stacking, such as stratified cross-validation and model cloning, can increase computational complexity and time, especially with large datasets or when a large number of base models are used.
- Dependence on Base Model Selection: The performance of Redesignate Stacking is still reliant on the selection of appropriate base models. If the base models themselves are weak or poorly suited to the problem, the meta-classifier may not perform as well, even with robust training techniques.
- Overfitting in Highly Complex Datasets: Despite its enhancements, Redesignate Stacking may still suffer from overfitting if the base models are too complex or if the dataset contains too many features. Ensuring that the base models are well-regularized is essential for maintaining the robustness of the classifier.

5.6 Practical Implications:

The **Redesignate Stacking Classifier** offers significant benefits for tasks such as **healthcare diagnostics**, **fraud detection**, and **financial forecasting**. Its **robustness** and **generalization** make it particularly valuable for datasets with imbalances or noisy labels.

Future Directions: Investigating integration with pre-trained models for specific domains like image recognition or natural language processing.Real-Time Processing: Optimizing the Redesignate Stacking framework for use in streaming data environments.Scalability: Exploring scalability for large-scale applications with distributed systems or cloud computing.

6. CONCLUSION

The Redesignate Stacking Classifier demonstrates superior performance compared to the Standard Stacking Classifier in terms of ROC-AUC**,** confusion matrix metrics**,** and generalization. While the improvements in accuracy and ROC-AUC are modest, the Redesignate Stacking method provides notable advantages in handling imbalanced data, robustness to noise, and model generalization. The statistical tests confirm that the performance differences are not statistically significant in terms of accuracy and misclassification rates, but the model offers practical advantages in real-world applications where generalization and model diversity are crucial.

This structured approach highlights how each classifier handles training and prediction, emphasizing differences in model handling and data aggregation. Both classifiers aim to improve prediction performance by combining multiple base models. The key difference is in the approach to ensuring the independence of each base model during cross-validation and prediction, which is more rigorously handled in the Redesignate Stacking Classifier through model cloning.

Further research could explore the application of Redesignate Stacking to more complex datasets, as well as the integration of other advanced techniques such as feature selection and hyperparameter optimization to further enhance performance.

REFERNCE

1. Abdollahi J, Nouri-Moghaddam B. Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction. Iran J Comput Sci. 2022;5(3):205–20. doi: 10.1007/s42044-022-00100-1. Epub 2022 Mar 21. PMCID: PMC8935256.
2. Aditi Gavhane; GouthamiKokkula; Isha Pandya; Kailas Devadkar, Prediction of heart disease using machine learning, in: 2018 Second International Conference on Electronics, Communication and Aerospace Technology, ICECA, 2018, pp. 1275–1278, http: //dx.doi.org/10.1109/ICECA.2018.8474922.
3. Ammar Mohammed, Rania Kora," A comprehensive review on ensemble deep learning: Opportunities and challenges, Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 2, February 2023, Pages 757-774, doi: https://doi.org/10.1016/j.jksuci.2023.01.014
4. B. Babenko, M. -H. Yang and S. Belongie, "A family of online boosting algorithms," 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, Japan, 2009, pp. 1346-1353, doi: 10.1109/ICCVW.2009.5457453.
5. Chih-Fong Tsai, Yu-Feng Hsu, A Meta-learning Framework for Bankruptcy Prediction, March 2013, Volume 32, Issue 2, Pages 167-179 ,https://doi.org/10.1002/for.1264
6. Czarnowski and P. Jędrzejowicz, "Stacking and rotation-based technique for machine learning classification with data reduction," 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, Poland, 2017, pp. 55-60, doi: 10.1109/INISTA.2017.8001132.
7. Czarnowski and P. Jędrzejowicz, "Stacking and rotation-based technique for machine learning classification with data reduction," 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Gdynia, Poland, 2017, pp. 55-60, doi: 10.1109/INISTA.2017.8001132.
8. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," in IEEE Access, vol. 10, pp. 99129-99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
9. Eun Sung Lee, Exploring the Performance of Stacking Classifier to Predict Depression Among the Elderly, IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, USA 23-26 August 2017, DOI: 10.1109/ICHI.2017.95
10. J Thomas, R.T. Princy, Human heart disease prediction system using data mining techniques, in: 2016 International Conference on Circuit, Power and Computing Technologies, ICCPCT, 2016, pp. 1–5, http://dx.doi.org/10.1109/ICCPCT.2016. 7530265.
11. J. Feng et al., "Effective techniques for intelligent cardiotocography interpretation using XGB-RF feature selection and stacking fusion," 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Houston, TX, USA, 2021, pp. 2667-2673, doi: 10.1109/BIBM52615.2021.9669694.
12. J. Niyogisubizo, L. Liao, Y. Lin, L. Luo, E. Nziyumva and E. Murwanashyaka, "A Novel Stacking Framework Based On Hybrid of Gradient Boosting-Adaptive Boosting-Multilayer Perceptron for Crash Injury Severity Prediction and Analysis," 2021 IEEE 4th International Conference on Electronics and Communication Engineering (ICECE), Xi'an, China, 2021, pp. 352-356, doi: 10.1109/ICECE54449.2021.9674567
13. K Shailaja, B Seetharamulu, M.A. Jabbar, Machine learning in healthcare: A review, in: 2018 Second International Conference on Electronics, Communication and Aerospace Technology, ICECA, 2018, pp. 910–914, http://dx.doi.org/10. 1109/ICECA.2018.8474918.
14. K. Sathyaki and R. Paily, "Leakage Reduction by Modified Stacking and Optimum ISO Input Loading in CMOS Devices," 15th International Conference on Advanced Computing and Communications (ADCOM 2007), Guwahati, India, 2007, pp. 220-227, doi: 10.1109/ADCOM.2007.85.
15. M. S. Bin Alam, M. J. A. Patwary and M. Hassan, "Birth Mode Prediction Using Bagging Ensemble Classifier: A Case Study of Bangladesh," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2021, pp. 95-99, doi: 10.1109/ICICT4SD50815.2021.9396909.
16. N. C. Oza, "Online bagging and boosting," 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, USA, 2005, pp. 2340-2345 Vol. 3, doi: 10.1109/ICSMC.2005.1571498.

17. S. Kumar, P. Kaur, and A. Gosain**.** "A Comprehensive Survey on Ensemble Methods," *2022 IEEE 7th International Conference for Convergence in Technology (I2CT)*, Mumbai, India, 2022, pp. 1-7. doi: 10.1109/I2CT54291.2022.9825269.
18. R. Pelossof, M. Jones, I. Vovsha and C. Rudin, "Online coordinate boosting," 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, Japan, 2009, pp. 1354-1361, doi: 10.1109/ICCVW.2009.5457454.
19. Ravinder Ahuja, Subhash C. Sharma, Maaruf Ali, A Diabetic Disease Prediction Model Based on Classification Algorithms, International Association of Educators and Researchers (IAER), 2019, pp. 44-52, Vol. 3, No. 3, DOI: 10.33166/AETiC.2019.03.005
20. Roberto Corizzo& Jacob Rosen, Stock market prediction with time series data and news headlines: a stacking ensemble approach, Journal of Intelligent Information Systems, 2023, Volume 62, pages 27–56, doi: https://doi.org/10.1007/s10844-023-00804-1
21. S. Kumar, P. Kaur and A. Gosain, "A Comprehensive Survey on Ensemble Methods," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-7, doi: 10.1109/I2CT54291.2022.9825269.
22. S. Wu, S. Deng and J. Cao, "A Three-dimension Stacking Model with Modified Genetic Algorithm," 2022 IEEE International Symposium on Product Compliance Engineering - Asia (ISPCE-ASIA), Guangzhou, Guangdong Province, China, 2022, pp. 1-6, doi: 10.1109/ISPCE-ASIA57917.2022.9971062.
23. Sikora, Riyaz and Al-laymoun, O'laHmoud (2014) "A Modified Stacking Ensemble Machine Learning, Algorithm Using Genetic Algorithms," Journal of International Technology and Information Management:Vol. 23: Iss. 1, Article 1.DOI: https://doi.org/10.58729/1941-6679.1061, Available at: https://scholarworks.lib.csusb.edu/jitim/vol23/iss1/1
24. T. Jiang, J. P. Li, A. U. Haq, A. Saboor and A. Ali, "A Novel Stacking Approach for Accurate Detection of Fake News," in IEEE Access, vol. 9, pp. 22626-22639, 2021, doi: 10.1109/ACCESS.2021.3056079.
25. T. Toharudin et al., "Boosting Algorithm to Handle Unbalanced Classification of PM2.5 Concentration Levels by Observing Meteorological Parameters in Jakarta-Indonesia Using AdaBoost, XGBoost, CatBoost, and LightGBM," in IEEE Access, vol. 11, pp. 35680-35696, 2023, doi: 10.1109/ACCESS.2023.3265019
26. U. Giridharan, V. Janardhanan, P. Ravi and B. V. Hiremath, "Classification of Fundus Images Using Modified Stacking Ensemble," 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 2022, pp. 330-335, doi: 10.1109/I4C57141.2022.10057942.
27. Urja Giridharan; Vedya Janardhanan; Prabha Ravi; Basavaraj V Hiremath, Classification of Fundus Images Using Modified Stacking Ensemble 4th International Conference on Circuits, Control, Communication and Computing (I4C), Bengaluru, 21-23 December 2022, DOI: 10.1109/I4C57141.2022.10057942
28. X. -D. Zeng, S. Chao and F. Wong, "Optimization of bagging classifiers based on SBCB algorithm," 2010 International Conference on Machine Learning and Cybernetics, Qingdao, China, 2010, pp. 262-267, doi: 10.1109/ICMLC.2010.5581054.
29. Y. Li and Z. Ye, "Boosting Independent Component Analysis," in IEEE Signal Processing Letters, vol. 29, pp. 1367-1371, 2022, doi: 10.1109/LSP.2022.3180680.
30. Yunpeng Zhao, DimitriosGouliasa and SetareSaremib, Enhancing Stacking Ensemble Performance Using Feature Engineering Techniques, 2023, Computers and Concrete, Vol. 32, No. 3, pp 233-246 https://doi.org/10.12989/cac.2023.32.3.233.