**Indonesian Journal of Electrical Engineering and Computer Science**
Vol. 36, No. 3, December 2024, pp. 1683∼1693
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v36.i3.pp1683-1693


# Stacking-based ensemble learning for identifying artist signatures on paintings

**Shintami Chusnul Hidayati, Agustinus Aldi Irawan Rahardja, Nanik Suciati**
Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

**Article Info**

***Article history:***

Received Jul 4, 2024
Revised Aug 7, 2024
Accepted Aug 25, 2024

***Keywords:***

Artist signature identification
Convolutional neural networks
Cultural heritage
Ensemble learning
Image classification

**ABSTRACT**

Identifying artist signatures on paintings is essential for authenticating artworks and advancing digital humanities. An artist's signature is a consistent element included in each painting that the artist creates, providing a unique identifier for their work. Traditional methods that rely on expert analysis and manual comparison are time-consuming and are prone to human error. Although convolutional neural networks (CNNs) have shown promise in automating this process, existing single-model approaches struggle with the diversity and complexity of artistic styles, leading to limitations in their performance and generalizability. Therefore, this study proposes an ensemble learning approach that integrates the predictive power of multiple CNN-based models. The proposed framework leverages the strengths of three state-of-the-art CNNs: EfficientNetB4, ResNet-50, and Xception. These models were independently trained, and the predictions were combined using a meta-learning strategy. To address class imbalance, data augmentation techniques and weighted loss functions were employed. The experimental results obtained on a dataset of more than 8,000 paintings from 50 artists demonstrate significant improvements over individual CNN architectures and other ensemble methods, thereby effectively capturing complex features and improving generalizability.



***Corresponding Author:***

Shintami Chusnul Hidayati
Department of Informatics, Institut Teknologi Sepuluh Nopember
ITS Sukolilo Campus, Surabaya 60111, Indonesia
Email: shintami@its.ac.id

## 1. INTRODUCTION

Identifying artist signatures on paintings is crucial for art authentication, origin research, and digital humanities [1]. Artist signatures provide essential information about the origin and authenticity of artworks, making their accurate identification indispensable for art historians, curators, and collectors. Traditional signature identification methods primarily rely on expert analysis and manual comparison, which are time-consuming and prone to human error [2], [3]. In recent years, advances in machine learning, particularly convolutional neural networks (CNNs), have demonstrated significant potential for automating this process [4]. CNNs excel at image recognition and classification tasks due to their ability to learn hierarchical feature representations [5]-[7]. However, current state-of-the-art methods, which typically rely on single CNN architectures, often face challenges due to the diversity and complexity of artistic styles, which limits their performance and generalizability [8]-[10].

Early studies have explored using handcrafted features for artist identification with a limited number of painter classess. For example, the study by Shamir *et al.* [11] assessed the use of fisher scores.

Condorovici *et al.* [12] integrated low-level features (i.e., 3D RGB Histograms and Gabor Energy Features) with classical machine learning techniques. Similarly, Alyannezhadi [13] extracted 11 statistical features, including edge detection and color histograms, and used hierarchical support vector machines (SVM) to enhance artist identification. Sheng and Jiang [14] applied handcrafted features to classify Chinese ink paintings. However, handcrafted features often require domain expertise and may not capture the full range of variability in artistic styles [15]-[17]. Handcrafted features are less robust to lighting, resolution, and noise variations, which are common in real-world datasets. In addition, they are typically less flexible and scalable than features learned by deep learning models; thus, they are less applicable to diverse and complex datasets.

Several studies have explored various approaches to improve the accuracy of signature identification. For example, Tan *et al.* [18] utilized end-to-end deep convolutional models to classify artists based on their painting styles and demonstrated improved classification accuracy compared to traditional methods that employed low-level descriptors. Saleh and Elgammal [19] combined handcrafted and CNN-based features to emphasize the importance of learning the correct feature extraction metric. Another study by Hong and Kim [20] developed a hybrid model for artist identification that combined hand-crafted local features with learned features. Vinayavekhin *et al.* [21], employed Siamese neural networks and a self-supervised learning approach to identify relationships paintings of Japanese and Western artists. Recently, some studies [4], [22], [23] further advanced this task by fine-tuning CNNs specifically for artist attribution. Their work highlights the effectiveness of transfer learning in adapting pretrained models to the artistic domain.

Despite these advancements, several challenges still need to be addressed regarding the use of conventional deep-learning methods for artist identification. Art datasets often suffer from class imbalance, where some artists represent significantly more work than others, leading to biased models. The wide range of styles and techniques used by different artists further complicates the task, as single-model approaches may not fully capture the intricate details needed for accurate identification. In addition, the limited availability of high-quality annotated data poses challenges, which makes it difficult to train deep learning models effectively. Overfitting remains a concern because models trained on small datasets may perform well on training data but must be generalized to unseen artworks.

Motivated by the limitations of existing methods, we propose a stacking ensemble approach to enhance the performance of artist signature identification. Stacking ensembles combine the strengths of multiple models, leveraging their diverse predictions to improve overall accuracy and robustness [24], [25]. By integrating several base learners with unique advantages, this approach can better capture subtle and complex features of artist signatures, mitigating individual model weaknesses and enhancing their strengths. Specifically, our method integrates the predictive power of multiple CNN-based models: EfficientNetB4 [26], ResNet-50 [27], and Xception [28]. These models are independently fine-tuned on a dataset of over 8,000 paintings from 50 different painters and their predictions are combined through a meta-learning strategy. To address class imbalance in the dataset, we employ data augmentation techniques such as random rotations, flips, and color adjustments, along with weighted loss functions during training to assign greater importance to less represented classes, thereby reducing bias toward more prevalent classes.

In summary, this study makes several key contributions. First, we developed an effective preprocessing pipeline to standardize and augment input images. Second, we integrated diverse CNN architectures to capture a wide range of artistic features. Third, we implemented a meta-learner to combine the predictions of the base models, thereby enhancing overall performance. These contributions address the limitations of existing methods and improve the automated identification of artist signatures.

The remainder of this paper is organized as follows. Section 2 details the methodology, including the preprocessing pipeline, base learner architectures, and ensemble learning approach. Section 3 describes the experimental setup, evaluation metrics, and experimental results. Finally, section 4 concludes the paper, summarizing the key contributions and outcomes of this study.

## 2. METHOD

This section details the proposed methodology, which integrates the predictive power of multiple CNN-based models to improve overall predictive performance. Given an input image, the proposed method automatically identifies the corresponding painter artists. Figure 1 illustrates the framework, which comprises three main steps: image preprocessing, base learning, and ensemble learning.

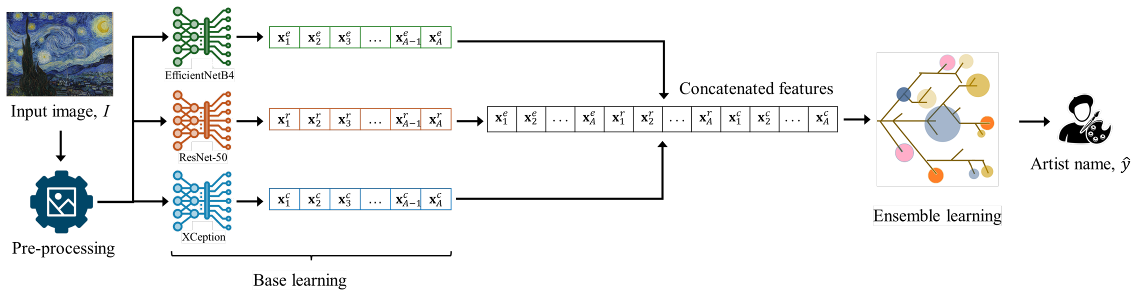



Figure 1. The proposed framework

### 2.1. Image preprocessing

In this step, the input images are standardized and augmented to ensure consistency for training and inference. Initially, the images are resized to specific dimensions to meet the unique architectural requirements of each base model. For ResNet-50, images are resized to 224 × 224 pixels. Xception, which requires larger input dimensions, utilizes images resized to 299 × 299 pixels. EfficientNetB4 benefits from even higher resolution inputs, with images resized to 380 × 380 pixels. In the following resizing step, color normalization is performed to standardize the input images. Each pixel value is scaled to fall within the range of 0 to 1 by dividing by the maximum pixel value, 255. Furthermore, pixel values are normalized to have zero mean and unit variance based on the dataset's mean and standard deviation. This normalization step accelerates the convergence of the training process by ensuring that the input data have a uniform distribution.

Several data augmentation techniques are applied to augment the dataset to prevent overfitting [29], [30]. These techniques increase the diversity of the training dataset. The random transformations used in this study include random rotations, horizontal and vertical flipping, and brightness and contrast adjustments. Random rotations within the range of -15° to +15° introduce variations in image orientation, thereby improving model invariance under slight rotations. Horizontal and vertical flipping simulate different perspectives, ensuring that the model recognizes artist signatures regardless of orientation. Additionally, random brightness and contrast adjustments are applied to simulate variations in the lighting conditions, thereby making the model more robust to different lighting environments.

### 2.2. Base learning

As mentioned previously, the ensemble learning approach harnesses the strengths of three state-of-the-art CNNs: EfficientNetB4 [26], ResNet-50 [27], and Xception [28]. Each base learner offers unique advantages, enabling the model to capture and analyze subtle and complex characteristics of an artist's paintings. Independently trained on a dataset of paintings, these base learners generate class predictions that collectively enhance the ensemble's overall performance.

The motivations for employing these three base learners are as follows. EfficientNetB4 is known for its optimized balance between accuracy and computational efficiency. This model employs a compound scaling method that uniformly scales the network depth, width, and resolution. This scaling approach enabled the model to achieve higher accuracy with fewer computational resources than traditional scaling methods. The ability to handle high-resolution inputs and its efficient use of parameters makes EfficientNetB4 a valuable component in our ensemble. ResNet-50, the second base learner in our ensemble, is a deep residual network renowned for its ability to address the vanishing gradient problem, which is common in deep neural networks. By using residual connections, ResNet-50 facilitates the training of very deep networks, allowing it to learn intricate patterns and features in images. Xception, which stands for "extreme inception", is another robust CNN architecture employed as a base learner. It builds upon the inception model using depth-wise separable convolutions. This architectural choice significantly reduces the number of parameters while maintaining high performance, enabling the model to learn spatial hierarchies in the input data efficiently. Xception's design allows it to capture detailed and complex features, making it well-suited for tasks involving fine-grained visual distinctions, such as identifying artist paintings.

### 2.3. Ensemble learning

This step integrates the classification results from the base learners-EfficientNetB4, ResNet-50, and Xception. In this integration, we first transform the classification results obtained from each base learner to one-hot encoding. In other words, each predicted class is represented by a vector whose length is equal to the number of classes in the dataset (50 classes of artists $A$ in this study). The predicted class is marked as 1 while all other classes are marked as 0. Once the one-hot encodings of the base learners are obtained, we concatenate them into a single feature vector. Let the feature vectors obtained from EfficientNetB4, ResNet-50, and Xception for image $I$ be represented as $\mathbf{x}_i^e$, $\mathbf{x}_i^r$, and $\mathbf{x}_i^c$, respectively, where each vector has 50 dimensions. Then, the concatenated feature vector for image $I$, $\mathbf{x}_i$, is given by (1):

$$\mathbf{x}_i = [\mathbf{x}_i^e, \mathbf{x}_i^r, \mathbf{x}_i^c]. \tag{1}$$

Therefore, the combined vector encapsulates the predictive information from the three base models, providing a rich and detailed representation of the input data.

The feature vector is then input to a decision-making process that employs ensemble methods. Suppose we have the dataset defined in (2):

$$D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, \tag{2}$$

where $\mathbf{x}_i$ represents the feature vector, $y_i$ is the target variable, and $N$ is the number of paintings in the dataset. The meta-learner builds $M$ decision trees, defined in (3):

$$\{T_m\}_{m=1}^{M}, \tag{3}$$

using bootstrap samples from the original dataset. A bootstrap sample is created by randomly selecting $N$ instances from the original dataset $D$ with replacement. Thus, some instances may be selected multiple times, while others may not be selected at all. Formally, let $D^*$ denote a bootstrap sample, as written in (4).

$$D^* = \{(\mathbf{x}_i^*, y_i^*)\}_{i=1}^{N}, \quad \text{where} \quad (\mathbf{x}_i^*, y_i^*) \in D. \tag{4}$$

Bootstrap sampling creates diverse training sets for each decision tree in the ensemble [31], [32]. This diversity helps reduce overfitting by ensuring that each tree is trained on a slightly different subset of the dataset, which captures various aspects of the dataset's underlying structure. Moreover, bootstrap sampling allows the model to estimate the prediction variance and improves the robustness and reliability of the final model.

Each tree $T_m$ in (2.3.) is constructed using a random subset of features at each split to enhance the diversity of the trees. For a given input $\mathbf{x}$, each tree $T_m$ provides a class prediction $h_m(\mathbf{x})$. The final prediction of the classifier is determined by aggregating the predictions of all the trees. Specifically, the output $\hat{y}$ is given by the mode of the predicted classes, as formulated in (5):

$$\hat{y} = \text{mode}\{h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_M(\mathbf{x})\}. \tag{5}$$

this majority voting mechanism ensures that the meta-learner classifier reduces overfitting and improves generalizability compared to individual decision trees.

## 3. EXPERIMENTS

This section presents the dataset, experimental setup, evaluation metrics, and results obtained using the proposed ensemble learning approach. We evaluated the performance of the base learners and ensemble method on a painting dataset. The results highlight the effectiveness of the proposed method.

### 3.1. Dataset

The dataset used in this study was collected from art challenge under the category of famous painters. It comprises a total of 8,446 paintings by 50 different painters [33]. Each image was associated with a specific artist, providing a labeled dataset suitable for supervised learning to train and evaluate the artist identification models. As shown in Figure 2, which shows the image distribution across 50 painters, the dataset exhibits an imbalance in the number of paintings painted by each artist. This imbalance challenges learning algorithms because models trained on imbalanced data may be biased toward more represented classes.

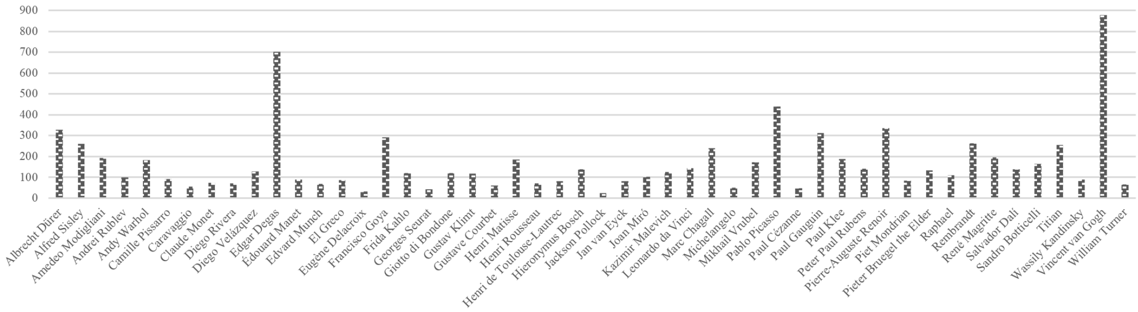



Figure 2. Number of paintings for each artist in the dataset (Better view in electronic version)

### 3.2. Baseline models

To evaluate the effectiveness of the proposed ensemble learning approach, we compared its performance to that of several baseline models. These baselines include individual CNN architectures and other ensemble methods. The individual CNN architectures considered are EfficientNetB4 [26], ResNet-50 [27], and Xception [28]. The ensemble methods include majority voting, weighted voting [34], and XGBoost [35].

As mentioned previously, one of the main reasons behind the success of ensemble methods is increasing the diversity among the base classifiers [36]. Specifically, in the majority voting approach, the majority class determines the final prediction that the base learners predict. Each base learner equally contributes to the final decision, and the class with the most votes is selected. The weighted voting method assigns different weights to predictions of each base learner based on individual accuracy performance. The final prediction is a weighted sum of the base learners' predictions, with higher-performing models significantly influencing the outcome. XGBoost is a more advanced ensemble method that uses gradient boosting to combine the predictions of multiple weak learners. XGBoost is known for its high accuracy and efficiency, making it a strong baseline for comparison.

### 3.3. Experimental setup

This subsection outlines the experimental setup. It includes the hardware and software configurations as well as the training procedures for the base learners. Additionally, it specifies the hyperparameters used for each model.

#### 3.3.1. Hardware and software configurations

The experiments were conducted on Google Collaboratory Pro, equipped with T4 graphics processing units (GPUs). The software environment included Python, TensorFlow, and Keras. These frameworks provide robust tools for implementing and training CNN architectures.

#### 3.3.2. Training procedures

Each base learner (EfficientNetB4, ResNet-50, and Xception) was independently trained on the paintings dataset. The dataset was divided into training, validation, and test sets with an 80-10-10 ratio. The training process involved fine-tuning the pretrained models on the ImageNet dataset to adapt them to the painter identification task. Each model was trained for 50 epochs using a batch size of 32, with early stopping based on validation loss to prevent overfitting. The Adam optimizer was used with an initial learning rate of 0.0001, which was reduced by a factor of 10 if validation loss did not improve over five consecutive epochs. For training the ensemble model, the meta-learner employed an ensemble of decision trees, with 100 trees constructed using bootstrap samples from the original dataset. Each tree was trained using a random subset of features at each split to enhance diversity.

### 3.4. Evaluation metrics

The models' performance was assessed using several metrics to provide a comprehensive understanding of their effectiveness in identifying painter signatures. The metrics included accuracy, precision, recall, and F1-score. In addition, accuracy and loss were evaluated during the training period.

Accuracy measures the overall correctness of the model by calculating the proportion of correctly classified instances out of the total number of instances. Precision, also known as positive predictive value, measures the accuracy of the positive predictions made by the model. High precision indicates a low false positive rate. Recall, or sensitivity, measures the model's ability to identify all relevant instances. High recall indicates that the model correctly identified the most positive instances; thus, high recall is necessary when the cost of false negatives is high. The F1-score is the harmonic mean of precision and recall, providing a metric that balances both concerns.

We plotted the accuracy and loss graphs for both the training and validation sets over epochs to visualize the training process and monitor overfitting and underfitting. The accuracy graph shows how the model's accuracy evolves during training, while the loss graph depicts the change in the loss function. These graphs help visualize the model's learning behavior and adjust the hyperparameters to improve performance.

### 3.5. Results and discussion

Table 1 summarizes the performances of different models in identifying artist signatures on paintings. The results highlight the effectiveness of the proposed ensemble learning approach compared to individual CNN architectures and other ensemble methods. This section discusses the implications and insights derived from these findings.

Table 1. Performance report (in %)

| Method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| EfficientNetB4 [26] | 41.52 | 32.80 | 33.87 | 33.33 |
| Resnet-50 [27] | 54.67 | 40.71 | 39.41 | 40.05 |
| Xception [28] | 66.67 | 54.89 | 65.40 | 52.03 |
| Majority voting | 67.82 | 60.37 | 52.46 | 56.13 |
| Weighted voting [33] | 67.94 | 59.59 | 53.77 | 56.53 |
| XGBoost [34] | 87.20 | 93.24 | 89.46 | 91.31 |
| Ours | **88.12** | **93.47** | **90.06** | **91.73** |

Among the individual base learners, Xception achieved the highest accuracy of 66.67%, with a precision of 54.89%, recall of 65.40%, and an F1-score of 52.03%. ResNet-50 and EfficientNetB4, while still performing reasonably well, lagged with accuracy values of 54.67% and 41.52%, respectively. These results highlight the strengths and limitations of using single-model approaches for this task. Xception's higher performance can be attributed to its depthwise separable convolutions, which effectively capture detailed and complex painting features. However, the best-performing single model (Xception) was below the desired performance levels, which indicates the need for a more robust solution.

Figure 3 presents the training and validation accuracy and loss for each base learner, which provides insights into the learning behavior and performance of each model over training epochs. Overall, EfficientNetB4 demonstrated signs of overfitting, with a large gap between the training and validation metrics (Figures 3(a) and 3(d)), whereas ResNet-50 exhibited fluctuations in performance despite its efficient learning (Figures 3(b) and 3(e)), which suggests its sensitivity to training dynamics. In contrast, Xception demonstrated strong generalizability, maintaining closely aligned training and validation metrics (Figures 3(c) and 3(f)). These findings underscore the strengths and weaknesses of each base learner and underscore the importance of combining their predictions into an ensemble model. By leveraging the diverse learning behaviors and feature extraction capabilities of EfficientNetB4, ResNet-50, and Xception, the proposed ensemble approach aims to mitigate individual model weaknesses and enhance overall performance in identifying artist signatures.

Moreover, from Table 1, it is evident that majority voting and weighted voting outperformed individual base learners with accuracy values of 67.82% and 67.94%, respectively. These methods combine the strengths of multiple models, leading to more balanced and accurate predictions. However, their performance metrics indicate that although they enhance robustness, they still need to reach the highest accuracy and reliability required for precise artist identification.

XGBoost, a powerful gradient boosting algorithm, significantly outperformed individual models and basic ensemble methods, achieving accuracy of 87.20%, precision of 93.24%, recall of 89.46%, and F1-score of 91.31%. Compared to Xception, XGBoost demonstrated an increase in accuracy of over 20 percentage points, rising from 66.67% to 87.20%. The precision notably improved from 54.89% to 93.24%, while the recall increased from 65.40% to 89.46%, and the F1-score rose from 52.03% to 91.31%. These results demonstrate the potential of advanced ensemble techniques to improve classification performance. XGBoost's ability to handle imbalanced data and capture complex patterns via boosting contributed to its superior performance.

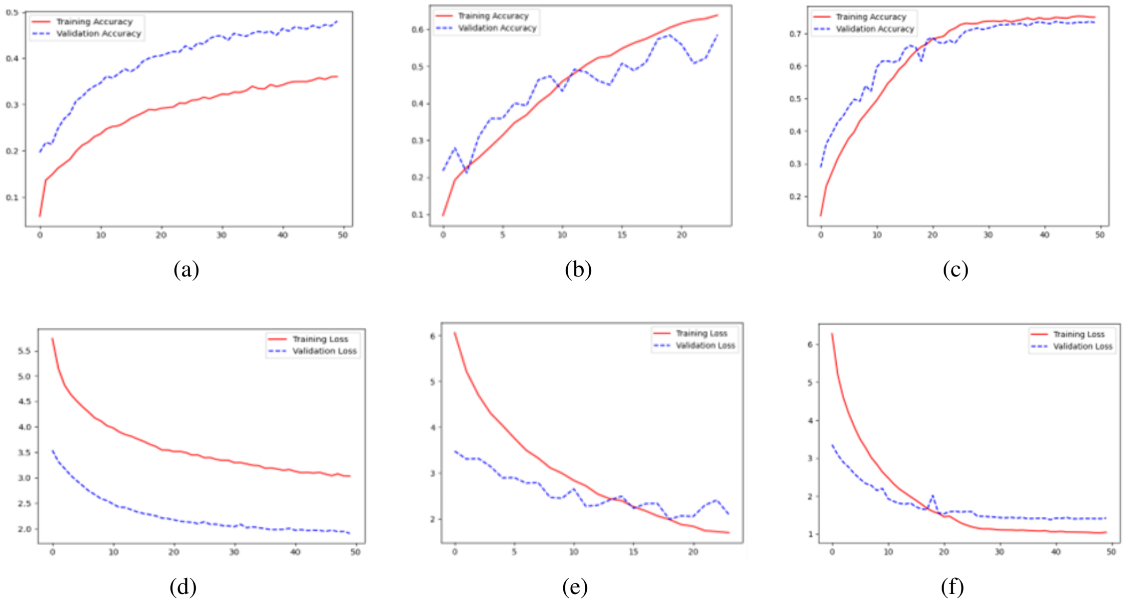


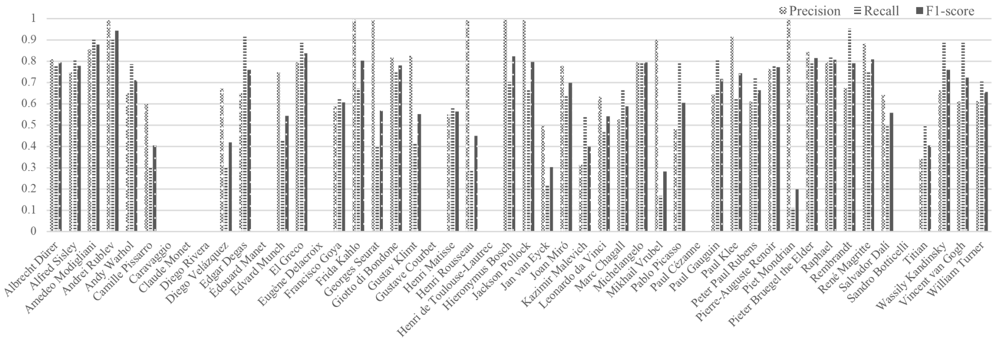

Figure 3. Training and validation accuracy; (a) EfficientNet accuracy, (b) ResNet-50 accuracy, and (c) Xception accuracy and loss; (d) EfficientNet loss, (e) ResNet-50 loss, and (f) Xception loss for each base learner

The proposed ensemble model, which integrates ResNet-50, Xception, and EfficientNetB4 via a stacking approach, achieved superior performance across all metrics. With an accuracy of 88.12%, precision of 93.47%, recall of 90.06%, and F1-score of 91.73%, our model not only outperformed individual base learners and simple ensemble methods but also improved upon the results of XGBoost, can be seen in Figure 4. As shown in Figures 4(a) and 4(b), both majority voting and weighted voting have limitations, particularly in identifying paintings belonging to minor artist classes, such as those by Caravaggio, Monet, and Diego Rivera. In contrast, XGBoost excelled in this task (Figure 4(c)), and the proposed method significantly enhanced the recognition of these minority classes (Figure 4(d)). Although the improvement over XGBoost was modest, it highlights the added value of our sophisticated ensemble method. These improvements underscore the substantial gains in performance achieved through the proposed ensemble approach, which effectively leverages the unique strengths of each base learner to capture diverse features and patterns that individual models may overlook.

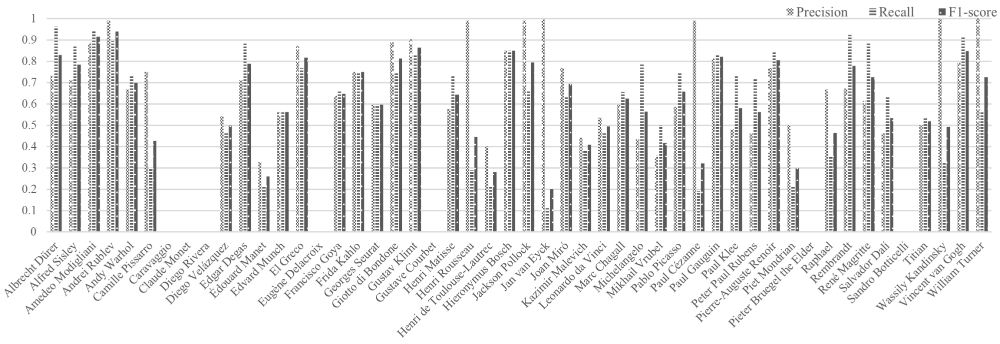
Figure 5 shows examples of paintings used in this study, along with their ground truth and predicted artist identifications by various models. In Figure 5(a), a painting by Andrei Rublev, most models incorrectly predicted the artist as Vincent van Gogh, Andy Warhol, or Giotto di Bondone, whereas the proposed model correctly identified Andrei Rublev. Figures 5(b) and 5(c) feature paintings by Albrecht Dürer. In Figure 5(b), predictions varied, with models suggesting artists like Rembrandt, Salvador Dalí, and Albrecht Dürer; however, the proposed model correctly identified Albrecht Dürer. Figure 5(c) showed higher accuracy, where most models correctly predicted Albrecht Dürer. This demonstrates the proposed model's effectiveness in accurately identifying artists, highlighting its robustness in handling diverse artistic styles.

Despite the superior performance of the proposed ensemble model, several challenges remain. Class imbalance in the dataset can lead to biased models favoring more represented classes. Techniques such as data augmentation and weighted loss functions were employed to mitigate this issue, further refinement is required. Additionally, the complexity of artistic styles requires models that learn a broad spectrum of features, which can be computationally intensive.

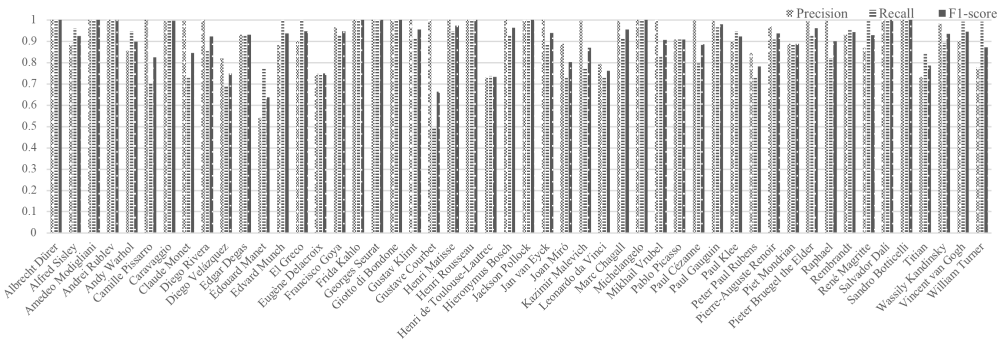
Future work will address these challenges by exploring more advanced data augmentation techniques and incorporating additional base learners to capture more diverse features. We also plan to investigate other advanced ensemble methods and meta-learning strategies to enhance the model's generalizability and performance. By continuing to refine and expand our approach, we aim to develop more accurate and reliable automated artist signature identification models.



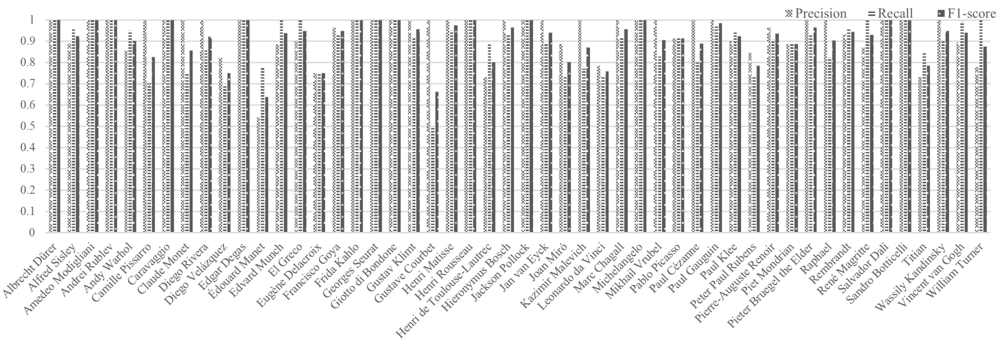

(a)




(b)




(c)




(d)

Figure 4. Performance results of the proposed approach compared to other ensemble methods; (a) ensemble by majority voting, (b) ensemble by weighted voting, (c) ensemble by XGBoost, and (d) pur proposed (Better view in electronic version)

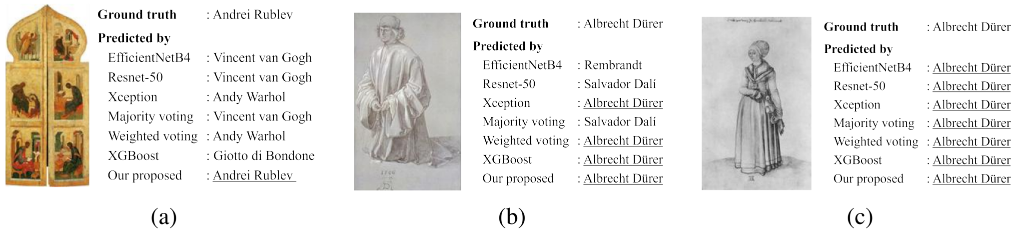



Figure 5. Examples of paintings in the study and their predicted artists; (a) a painting by Andrei Rublev, (b) a painting by Albrecht Dürer, and (c) a painting by Albrecht Dürer

## 4. CONCLUSION

This paper presented a novel ensemble learning approach for identifying artist signatures on paintings, leveraging the strengths of three state-of-the-art CNNs: EfficientNetB4, ResNet-50, and Xception. The proposed methodology integrated these diverse models through a sophisticated meta-learning strategy that combined predictions to enhance overall performance. Through comprehensive image preprocessing, including resizing, normalization, and data augmentation, we ensured that the input data was consistent for training and inference. The use of a meta-learner with multiple decision trees further refined the predictions, reducing overfitting and improving generalizability across the test data. Experimental results on a dataset of over 8,000 paintings from 50 artists demonstrated that the proposed model outperformed individual CNN architectures and other ensemble methods, such as majority voting, weighted voting, and XGBoost.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2024.

## REFERENCES

[1] J. Wilmerding, “Signs of the artist: signatures and self-expression in American paintings,” *Choice Reviews Online*, vol. 41, no. 07, pp. 41-3861-41–3861, Mar. 2004, doi: 10.5860/CHOICE.41-3861.

[2] S. Lyu, D. Rockmore, and H. Farid, “A digital technique for art authentication,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 49, pp. 17006–17010, Dec. 2004, doi: 10.1073/pnas.0406398101.

[3] G. Polatkan, S. Jafarpour, A. Brasoveanu, S. Hughes, and I. Daubechies, “Detection of forgery in paintings using supervised learning,” in *Proceedings - International Conference on Image Processing, ICIP*, Nov. 2009, pp. 2921–2924, doi: 10.1109/ICIP.2009.5413338.

[4] C. Comert, M. Ozbayoglu, and C. Kasnakoglu, “Painter prediction from artworks with transfer learning,” in *2021 7th International Conference on Mechatronics and Robotics Engineering, ICMRE 2021*, Feb. 2021, pp. 204–208, doi: 10.1109/ICMRE51691.2021.9384828.

[5] Z. M. Liu, Y. Y. Chen, S. Hidayati, S. C. Chien, F. C. Chang, and K.-L. Hua, “3D model retrieval based on deep Autoencoder neural networks,” in *2017 International Conference on Signals and Systems (ICSigSys)*, May 2017, pp. 290–296, doi: 10.1109/ICSIGSYS.2017.7967059.

[6] S. C. Hidayati and Y. Anistyasari, “Body shape calculator: understanding the type of body shapes from anthropometric measurements,” in *ICMR 2021 - Proceedings of the 2021 International Conference on Multimedia Retrieval*, Aug. 2021, pp. 461–465, doi: 10.1145/3460426.3463582.

[7] Q. Qi, A. Ahmad, and W. Ke, “Image classification based on few-shot learning algorithms: a review,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 2, pp. 933–943, Aug. 2024, doi: 10.11591/ijeecs.v35.i2.pp933-943.

[8] L. Alzubaidi et al., “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[9] J. K. Lee *et al.*, “Resource-efficient convolutional networks: a survey on model-, arithmetic-, and implementation-level techniques,” *ACM Computing Surveys*, vol. 55, no. 13 s, pp. 1–36, Dec. 2023, doi: 10.1145/3587095.

[10] J. Valencia, G. G. Pineda, V. G. Pineda, A. Valencia-Arias, J. Arcila-Diaz, and R. T. de la Puente, “Using machine learning to predict artistic styles: an analysis of trends and the research agenda,” *Artificial Intelligence Review*, vol. 57, no. 5, p. 118, Apr. 2024, doi: 10.1007/s10462-024-10727-0.

[11] L. Shamir, T. Macura, N. Orlov, D. M. Eckley, and I. G. Goldberg, “Impressionism, expressionism, surrealism,” *ACM Transactions on Applied Perception*, vol. 7, no. 2, pp. 1–17, Feb. 2010, doi: 10.1145/1670671.1670672.

[12] R. Condorovici, C. Florea, and C. Vertan, “Author identification for digitized paintings collections,” in *ISSCS 2013 - International Symposium on Signals, Circuits and Systems*, Jul. 2013, pp. 1–4, doi: 10.1109/ISSCS.2013.6651197.

[13] M. M. Alyannezhadi, H. Dabbaghan, S. Moghani, and M. Forghani, "A painting artist recognition system based on image processing and hierarchical SVM," in *2019 IEEE 5th Conference on Knowledge Based Engineering and Innovation, KBEI 2019*, Feb. 2019, pp. 537–541, doi: 10.1109/KBEI.2019.8734911.

[14] J. Sheng and J. Jiang, "Style-based classification of Chinese ink and wash paintings," *Optical Engineering*, vol. 52, no. 9, p. 093101, Sep. 2013, doi: 10.1117/1.oe.52.9.093101.

[15] T. Georgiou, Y. Liu, W. Chen, and M. Lew, "A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision," *International Journal of Multimedia Information Retrieval*, vol. 9, no. 3, pp. 135–170, Sep. 2020, doi: 10.1007/s13735-019-00183-w.

[16] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: a survey," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 23–79, Jan. 2021, doi: 10.1007/s11263-020-01359-2.

[17] R. Yelchuri, J. K. Dash, P. Singh, A. Mahapatro, and S. Panigrahi, "Exploiting deep and hand-crafted features for texture image retrieval using class membership," *Pattern Recognition Letters*, vol. 160, pp. 163–171, Aug. 2022, doi: 10.1016/j.patrec.2022.06.017.

[18] W. R. Tan, C. S. Chan, H. E. Aguirre, and K. Tanaka, "Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification," in *Proceedings - International Conference on Image Processing, ICIP*, Sep. 2016, vol. 2016-August, pp. 3703–3707, doi: 10.1109/ICIP.2016.7533051.

[19] B. Saleh and A. Elgammal, "Large-scale classification of fine-art paintings: learning the right metric on the right feature," *arXiv preprint*, 2015, doi: 10.11588/dah.2016.2.23376.

[20] Y. Hong and J. Kim, "Art painting detection and identification based on deep learning and image local features," *Multimedia Tools and Applications*, vol. 78, no. 6, pp. 6513–6528, Mar. 2019, doi: 10.1007/s11042-018-6387-5.

[21] P. Vinayavekhin, B. Khomkham, V. Suppakitpaisarn, P. Codognet, T. Terada, and A. Miura, "Identifying relationships and classifying western-style paintings: machine learning approaches for artworks by western artists and Meiji-era Japanese artists," *Journal on Computing and Cultural Heritage*, vol. 17, no. 1, pp. 1–18, Feb. 2024, doi: 10.1145/3631136.

[22] E. Cetinic, T. Lipic, and S. Grgic, "Fine-tuning convolutional neural networks for fine art classification," *Expert Systems with Applications*, vol. 114, pp. 107–118, Dec. 2018, doi: 10.1016/j.eswa.2018.07.026.

[23] H. Ugail, D. G. Stork, H. Edwards, S. C. Seward, and C. Brooke, "Deep transfer learning for visual analysis and attribution of paintings by Raphael," Heritage Science, vol. 11, no. 1, p. 268, Dec. 2023, doi: 10.1186/s40494-023-01094-0.

[24] N. R. Octavianto and A. Wibowo, "Stacking classifier method for prediction of human body performance," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 34, no. 3, pp. 1832–1839, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp1832-1839.

[25] L. Bagdadi and B. Messabih, "Distributed denial of service attacks classification system using features selection and ensemble techniques," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 34, no. 3, pp. 1868–1878, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp1868-1878.

[26] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, vol. 2016-Decem, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[28] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Jul. 2017, vol. 2017-January, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.

[29] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[30] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognition*, vol. 137, p. 109347, May 2023, doi: 10.1016/j.patcog.2023.109347.

[31] R. Duangsoithong and T. Windeatt, "Bootstrap feature selection for ensemble classifiers," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6171 LNAI, P. Perner, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 28–41.

[32] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, Feb. 2023, doi: 10.1016/j.jksuci.2023.01.014.

[33] "Art challenge." https://artchallenge.ru/ (accessed: 2023-06-01).

[34] A. Dogan and D. Birant, "A weighted majority voting ensemble approach for classification," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, Sep. 2019, pp. 1–6, doi: 10.1109/UBMK.2019.8907028.

[35] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 13-17-August-2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[36] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: a review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/j.engappai.2022.105151.

## BIOGRAPHIES OF AUTHORS

**Shintami Chusnul Hidayati** received her B.S. degree in Informatics from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, and her M.S. and Ph.D. degrees in Computer Science and Information Engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan. She is currently an Assistant Professor in the Department of Informatics at ITS. Prior to joining ITS, she was a Postdoctoral Research Associate at the Research Center for Information Technology Innovation, Academia Sinica, Taipei, from 2017 to 2019. Her current research interests include machine learning, data mining, and their applications in multimedia analysis, information retrieval, and computer vision. She has received several prestigious research awards and scholarships. She is also an honorary member of Phi Tau Phi. She can be contacted at email: shintami@its.ac.id.

**Agustinus Aldi Irawan Rahardja** received his Bachelor's degree in Informatics from Institut Teknologi Sepuluh Nopember (ITS), Indonesia, in 2024. He was also an exchange student at Asia University, Taiwan, in 2021. He can be contacted at email: 05111942000010@student.its.ac.id.

**Nanik Suciati** received the master's degree in computer science from the University of Indonesia, in 1998, and the Dr.Eng. degree in information engineering from Hiroshima University, in 2010. She is currently a Professor with the Department of Informatics, Institut Teknologi Sepuluh Nopember. She has published more than 50 journal articles and conference papers on computer science. Her research interests include computer vision, computer graphics, and artificial intelligence. She can be contacted at email: nanik@its.ac.id.