2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

# Development of an Image Recognition Model Using an Image Search Function Based on Multiple Pre-Trained Models

Hirotada Kuragane and Takeshi Sasaki

***Abstract*—Machine learning is widely utilized for data analysis and decision-making, with supervised and unsupervised learning being the primary approaches. However, models suffer from overfitting, where they become overly adapted to the training data. To address this issue, semi-supervised learning has been employed. Semi-supervised learning is an effective technique for dealing with large datasets that are difficult to label, but it faces limitations in fields where ensuring data diversity and quantity is challenging. This paper proposes a robust image recognition model utilizing image search functions from Google. The proposed model improves accuracy by utilizing the order of search results to collect a variety of data and evaluating their reliability. In this paper, the order of search results is defined as "image search depth" to measure the correlation between reliability and accuracy. While it is easy to collect large amounts of data from Google through automated methods, there is a risk that unrelated data could be included, potentially affecting the model's accuracy. To address this issue, the model is trained with automated preprocessing. As part of this preprocessing, inference is performed on all images in the dataset using multiple pre-trained models that were trained on randomly selected images from the dataset to compute predictions. Images with the prediction above a certain threshold are selected as training data to enhance the final model ' s accuracy. To assess the contribution of preprocessing to accuracy improvement, we calculate accuracy by varying the number of parallel pre-trained models and the threshold values. Furthermore, the final model is evaluated using CIFAR-100 to objectively demonstrate its performance. The results indicate that image search depth does not contribute to model accuracy, while the number of parallel pre-trained models and the threshold significantly impact accuracy.**

## I. Introduction

Machine learning is widely used in modern society to extract data and patterns, and to make predictions and decisions.Supervised learning and unsupervised learning are major approaches in data analysis and pattern recognition within the field of machine learning [1][2][3]. The accuracy of predictions made by a learning model is highly dependent on the quality and quantity of the data used for training. If the data are biased, the model may overfit the training data, which exacerbates the problem of poor generalization performance for unknown data, a phenomenon known as overfitting [4][5]. To solve this problem, a large amount of data must be available.

Hirotada Kuragane and Takeshi Sasaki are with the Graduate School of Shibaura Institute of Technology, 3-7-5 Toyosu, Koto-ku, Tokyo, Japan. cy20231@shibaura-it.ac.jp sasaki-t@shibaura-it.ac.jp

However, the disadvantage is that the more data to be prepared, the more time required for labeling. Therefore, semi-supervised learning is currently used. Semi-supervised learning is a powerful technique for handling large data sets that are difficult to label, and classifiers based on semi-supervised learning and hybrid generative-discriminative approaches are known to exhibit high classification performance [6][7][8]. Although this method is effective when there is accurate labeled data and a large amount of unlabeled data, maintaining recognition accuracy becomes challenging when new objects or concepts frequently emerge. Therefore, this paper proposes a robust image recognition model utilizing image search functionalities implemented by multiple search engines such as Google. Search results are returned in order of the strength of the relationship with the label; by leveraging this property, diverse and ample data with labels can be obtained. However, there is a risk that noisy data, which is unrelated to the label, may be mixed in, potentially reducing the model's accuracy [9][10]. This issue can be addressed by quantifying how reliable each image data is and setting a threshold to select which data to include in the final model construction. To quantify the reliability of the images, we employ ensemble learning. We train multiple types of pre-trained models and use bagging, followed by taking a simple average to reach a conclusion. Additionally, since image search functionalities on platforms like Google and Yahoo! return search results in order of the strength of the relationship with the label, there will be differences in accuracy between images that appear early in the results and those that appear later [11][12]. This paper defines the position of an image in the search results as "image search depth"" and seeks to determine the correlation between image search depth and accuracy. By weighting the quantified reliability of images according to image search depth, the paper aims to realize a robust image recognition model that takes image search depth into account. If this method is implemented in practice, it will enable the automatic detection of low-quality data. This is particularly valuable in situations where manually preparing a dataset with adequate quality and quantity is challenging. By identifying and addressing low-quality data, it will become possible to train highly accurate models, even when the dataset contains a significant proportion of low-quality data.

979-8-3315-3161-4/25/$31.00 ©2025 IEEE

## II. Related Studies

Tollari et al. [13] propose a method to enhance image classification accuracy by integrating visual features (such as color, shape, orientation, and edges) with textual information (keywords) in image search engines. Their "visual-text fusion" approach is applied to both news photos and landscape/animal images, where manually assigned keywords are used as text vectors, which are then combined with classification models based on these visual features. This method demonstrates a significant improvement in classification accuracy, with results showing up to a 50% increase compared to using textual information alone. Additionally, by filtering out irrelevant images from web search results using visual features, they effectively remove noise that does not match the search query. For this experiment, we propose a semi-supervised automatic preprocessing method that utilizes unlabeled images collected from Google. While their method relies on manual indexing, our approach aims to reduce the noise present in large datasets automatically gathered from web searches, leveraging the volume of images available online. Specifically, we employ multiple pre-trained models to process the collected images. The predictions from each model are aggregated to compute a confidence score for each image, allowing us to automatically filter out low-confidence images as noise. This approach introduces a novel method for automatically cleaning label data, leading to the development of a more accurate image classification model. D. Müller et al. [14] present a detailed analysis of optimizing ensemble learning techniques using deep convolutional neural networks (CNNs) for medical image classification. Their research investigates the effectiveness of combining multiple CNN models to enhance classification performance, with a specific focus on improving accuracy in medical image classification. The authors propose a method to optimize overall performance by complementing the strengths of individual models through combinations of different CNN architectures. Specifically, by merging the prediction results from multiple models, they achieve improvements in accuracy that cannot be attained with a single model alone. The paper also examines optimization techniques for ensemble learning, including model selection, weighting, and fusion strategies, and demonstrates their effectiveness through experiments. This study provides a practical approach to image classification using ensemble learning and offers valuable insights for improving classification accuracy, making it an important contribution in the context of this research.

## III. Proposed Method

We propose an automatic preprocessing method based on semi-supervised learning using unlabeled images collected from Google, as shown in Fig. 1. The goal of this method is to efficiently adapt to diverse datasets while reducing the burden of manual labeling. The process begins with the data collection phase, where a diverse set of images is gathered from Google using automated scraping techniques. Since directly training on these images might introduce irrelevant data not related to the search terms—potentially degrading the model's accuracy—we incorporate a strategy to automatically identify and exclude such irrelevant data. This is achieved by training multiple preprocessing models using the search terms of the collected images as labels. For each image, the prediction results from each pre-trained model are summed, averaged, and used to compute a confidence level. Images with low confidence are then classified as noise and removed from the dataset. The final model is trained on the cleaned dataset, resulting in a highly accurate image classification model. To elaborate further, the preprocessing begins by scraping a large volume of images from Google using specific search terms. These search terms serve as the initial labeling mechanism, assigning tentative labels to each image. However, since web-scraped images may include mislabeled or unrelated content, we introduce a filtering stage using multiple pre-trained models to evaluate the quality and relevance of each image. In the filtering stage, pre-trained models are created by training on random subsets of 70% of the collected images. This process is repeated ten times, generating ten models based on different training subsets. This diversity improves the robustness of the image evaluation. Once these models are established, they are applied to the entire dataset, and the prediction scores for each image are aggregated to compute an average confidence score. Images with consistently high confidence across multiple models are retained for further training, while those with low confidence are discarded as noise. This filtering method ensures that only high-quality data remains in the dataset, significantly reducing the time and effort typically required for manual labeling and filtering. After the noisy images are removed, the refined dataset is used to train the final image classification model. The clean dataset provides a more accurate representation and better labeling, ultimately leading to improved classification performance. Additionally, the relationship between search depth and image accuracy is an important consideration. It is known that the accuracy of images tends to decrease as the search depth increases in search engines like Google. To investigate this correlation and prevent low-confidence, irrelevant images from negatively impacting the results, we employ robust regression methods resistant to outliers. Wada et al. [15] estimate the impact of various weight functions and residual scales on robust regression. They evaluated Tukey ' s biweight function and Huber's weight function for the weight function, and average absolute deviation (AAD) and median absolute deviation (MAD) for scale adjustment constants. They concluded that the properties of the weight function should take precedence over estimation efficiency and computational efficiency. In this paper, to improve model accuracy, we use Tukey ' s biweight function to exclude extreme outliers.

$$w(e) = \begin{cases} \left[1 - \left(\frac{|e|}{c}\right)^2\right]^2 & \text{if } |e| \le c \\ 0 & \text{if } |e| > c \end{cases} \tag{1}$$

is adopted. Here $e$ and $c$ denote the the residual and the threshold, respectively. Additionally, while there was no

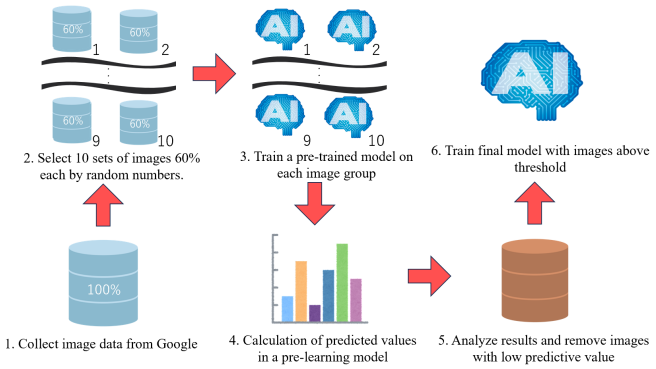


Fig. 1. Proposed Method

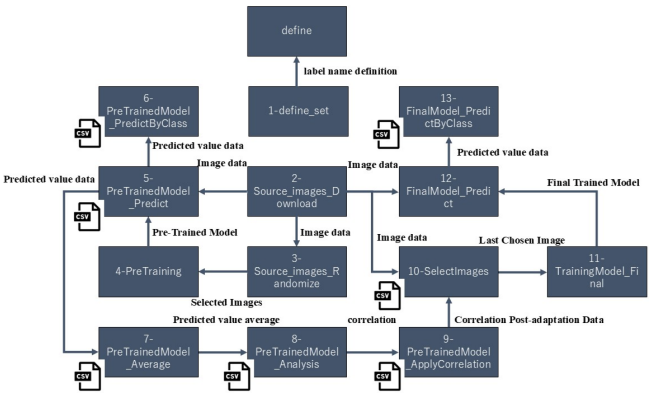


Fig. 2. Component Diagram

significant difference in estimation accuracy between the two scale parameter adjustment constants, they concluded that using the average absolute deviation (AAD) leads to faster convergence and lower computational costs. Therefore, this paper adopts

$$\hat{\sigma}_{\mathrm{AAD}} = \mathrm{mean}\left(|y_i - \mathrm{mean}(y_i)|\right) \tag{2}$$

where $y$ denotes the data set. By weighting the predicted values of images based on the correlation obtained in this manner, and by setting a threshold to determine which images to use for model construction, this paper attempts to build a robust image recognition model using image search functions.

## IV. Experiments

### A. Experimental Procedure and Workflow

This experiment follows the process outlined in Fig. 2. The steps are as follows:

1) Define_Set: Categories and items are configured and saved collectively in a file.
2) Source_Images_Download: HTTP requests are sent to Google Image Search to collect 500 images per item, which are stored in the data pool.The request query is ”category + item”.
3) Source_Images_Randomize: For each label, 70% (350 images) are randomly selected from the data pool and saved as training data for the pre-trained models.
4) PreTraining: The pre-training process is carried out with model settings specified in Table III.
5) PreTrainedModel_Predict: The pre-trained models estimate predictions for all images in the data pool.
6) PreTrainedModel_Predict_ByClass: Prediction values are classified by category and logged in CSV files.
7) PreTrainingModel_Average: The average prediction values for the collected images are calculated.
8) PreTrainingModel_Analysis: A regression analysis is performed to examine the relationship between image search depth and model accuracy.
9) PreTrainingModel_ApplyCorrelation: The computed correlation is applied as weights to the prediction values, refining the confidence score for each image.
10) SelectImages: The confidence values are approximated to a normal distribution, and only images with a probability above a set threshold are selected as training data for the final model.
11) TrainingModel_Final: The final model is trained using the same settings as the pre-trained models.
12) FinalModel_Predict: Prediction values are calculated for all images in the data pool.
13) FinalModelPredictByClass: Accuracy verification is conducted by classifying the predictions 3

Also I conducted two types of experiments. In the first experiment, I collected images of 10 categories and 10 items as Table I that I selected, and compared the accuracy of the final model with 10 pre-trained models that were created during the experiment without preprocessing. In the second experiment, I randomly selected 5 categories as Table II from the labels in CIFER-100, trained a model using the image search results as training data, and evaluated the accuracy of the model using the CIFER-100 dataset. As shown in Table III we chose ResNet18 for training the model. Because there is a concern of overfitting in the flow of learning the same image multiple times, we tried ResNet18 and ResNet50 and found no significant difference, so we chose ResNet18, which is faster to process. ResNet also has guaranteed accuracy for ImageNet.

### B. Evaluation experiments

This method aims to detect noisy data by estimating the predicted values of images in the data pool using multiple pre-trained models, thereby reduce the influence of inaccurate pre-trained models that have learned a lot of noisy data. Therefore, in order to investigate how the number of pre-trained models contributes to the accuracy of the final model, the number of pre-trained models is set from 1 to 10, and the final model is constructed for each pattern. The estimated predictions are used to select the teacher data for the final model by setting a threshold value. In order to determine how the threshold value contributes to the accuracy of the final model, the final model is constructed by setting the threshold value between 0% and 70% in increments of 5%. However, as the threshold is raised, the data used to evaluate the

TABLE I
LIST OF LABELS (EXPERIMENT1)

| Category | Items |
|---|---|
| World Heritage Sites | Angkor Wat, Canadian Rocky Mountain Parks, Grand Canyon, Great Barrier Reef, Sydney Opera House, Historic Centre of Florence, Machu Picchu, Mont Saint-Michel, Great Wall of China, Historic City of Ayutthaya |
| Geographical Features | Mountains, Plains, Rivers, Lakes, Coastlines, Deserts, Islands, Forests, Wetlands, Glaciers |
| Cuisine | Italian Cuisine, French Cuisine, Japanese Cuisine, Chinese Cuisine, Thai Cuisine, Indian Cuisine, Mexican Cuisine, Spanish Cuisine, Turkish Cuisine, American Cuisine |
| Dog Breeds | Labrador Retriever, Golden Retriever, Pomeranian, Dachshund, Siberian Husky, French Bulldog, Doberman Pinscher, Poodle, Border Collie, Schnauzer |
| Emotions | Joy, Sadness, Anger, Anxiety, Happiness, Jealousy, Fear, Relief, Tension, Excitement, Calm |
| Artistic Styles | Ancient Egyptian, Ancient Greek and Roman, Medieval European, Renaissance, Baroque, Rococo, Romanticism, Impressionism, Modernism, Postmodernism |
| Fashion Styles | Casual, Business Casual, Sporty, Vintage, Bohemian, Elegant, Rock, Preppy, Military, Romantic |
| Car Types | Sedan, SUV, Hatchback, Coupe, Convertible, Minivan, Truck, Crossover, Station Wagon, Sports Car |
| Hairstyles | Bob, Long Hair, Short Cut, Perm, Man Bun, Pixie Cut, Fade Cut, Long Layer, French Twist, Deep Side Part |
| Capitals | Cairo, New Delhi, Paris, Brasília, Berlin, Moscow, London, Washington D.C., Beijing, Tokyo |

TABLE II
LIST OF LABELS (EXPERIMENT2)

| Category | Items |
|---|---|
| fish | aquarium fish, flatfish, ray, shark, trout |
| flowers | orchids, poppies, roses, sunflowers, tulips |
| food containers | bottles, bowls, cans, cups, plates |
| Large man-made outdoor things | bridge, castle, house, road, skyscraper |
| Large natural outdoor scenes | cloud, forest, mountain, plain, sea |

model is also reduced and overlearning occurs, resulting in a significant decrease in accuracy for unknown data. Therefore, the second experiment using Table II is used. By estimating the images registered in CIFER-100 with the final model constructed by the above method and making a correct or incorrect decision, the accuracy can be estimated to unknown data.

### C. Result

First, we examined the correlation between depth of image retrieval and accuracy, and found that the accuracy of the pre-trained model and the predictive value of the data pool images varied significantly by item name, as shown in Fig. 2 and Fig. 3. When the accuracy of the pre-trained model itself was high and the predictive value of the data pool was also high, there was a negative correlation between rank and accuracy. On the other hand, for item names with low accuracy and low predictive values, there was a large variation in the prediction values for the data, and the relationship between rank and accuracy did not appear as a clear correlation. Next, Table I shows the results of comparing the accuracy

TABLE III
MODEL SETTINGS

| Model | ResNet18 |
|---|---|
| loss function | cross entropy loss |
| optimizer | stochastic gradient descent method |
| learning rate | 0.001 |
| moment function | 0.9 |
| epoch | 10 |
| test ratio | 0.3 |

of the pre-trained model and the final model trained with the labels in Table I. The threshold for the final model in this table is set at 5%. As shown in Table IV , Although there was a large difference in average accuracy by item name, the accuracy of the final model was greatly improved even within categories. The accuracy of models with and without applying the correlation between image search depth and accuracy was compared for each threshold, and the results are illustrated in Fig.5 and Fig.6. A common feature observed in both cases is that applying the correlation between image accuracy and rank to the prediction values of the data pool did not significantly improve the accuracy of the image models. However, raising the threshold to limit the images used as training data significantly improved accuracy. This approach, though, tends to select images that are mechanically easier to classify, which may lead to improved accuracy but potentially weaker performance as an image classification model. Additionally, reducing the amount of training data can negatively affect performance on unknown data, so simply increasing the threshold without a strategic approach is not advisable. Another aspect contributing to the differences in accuracy is the rate at which accuracy converges. Categories with initially high accuracy began to converge around a threshold of 40%, with minimal further improvement in accuracy afterward. In contrast, categories with initially low accuracy showed a monotonous increase in accuracy without convergence. Setting the threshold above 75% led to the creation of items with no training data, making it impractical to ensure accuracy using this method alone. Another notable feature is the difference in the rate of decrease in the number of images with increasing threshold. For the high accuracy category, setting the threshold to 50% yields 3,308 training images, whereas for the low accuracy to the application of the normal distribution approximation to the predictions and the threshold to the probability. Images in the high accuracy category have a smaller variance. Finally, Fig. 7 and Fig. 8 summarize the accuracy of the models learned in the categories corresponding to CIFAR-100, as well as the accuracy calculated by varying the number of parallelisms of the pre-trained models. Firstly, the results show that with a parallelism of 1, the pre-trained models tend to be biased, leading to very low accuracy. Furthermore, increasing the threshold does not improve the accuracy, indicating that even identifying images that are mechanically easier to classify is challenging. On the other hand, increasing the number of parallel pre-training models generally improves accuracy, particularly up to around 7 parallel models, where

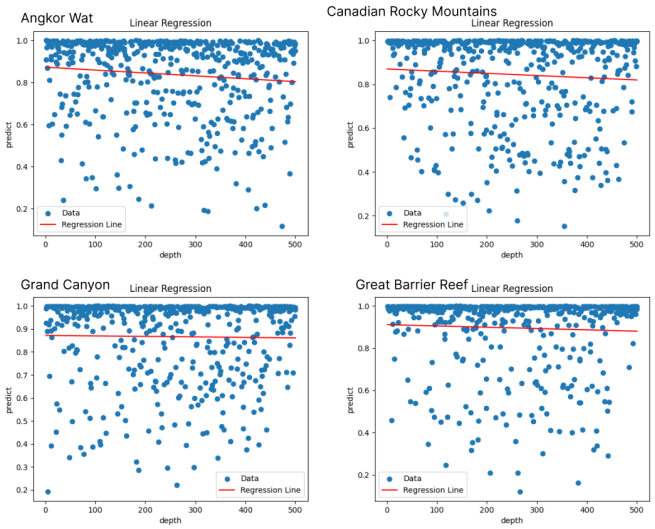


Fig. 3. Models with high accuracy and estimates (Category:World Heritage)

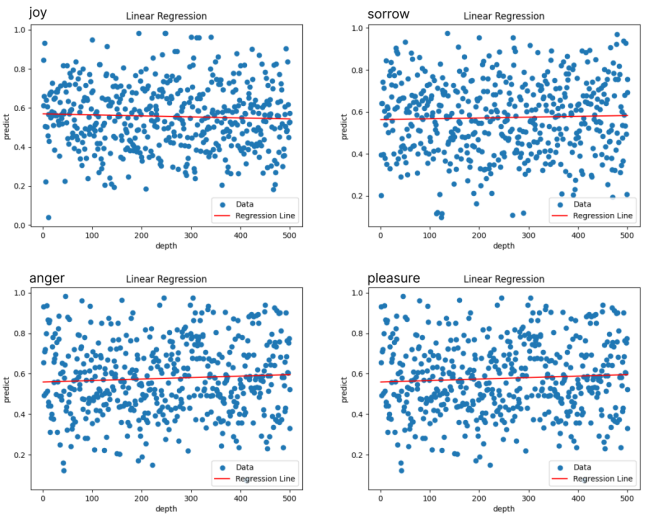


Fig. 4. Models with low accuracy and estimates (Category:color)

convergence in accuracy is observed. Increasing the number of models beyond 10 does not result in significant further improvement. Fig. 9 and Fig. 10 are regarding the graphs calculated using the CIFAR-100 dataset, no significant accuracy improvement is observed with increasing thresholds. Instead, accuracy sharply declines when the threshold is raised to 70%, regardless of the number of parallel models used. This suggests that the issue is independent of parallelism and is likely due to significant differences in image quality between the CIFAR-100 dataset images and those collected via Google Image Search. CIFAR-100 collects diverse images with varying colors and sizes within the same label, which might contribute to this discrepancy.

## V. Conclusion

This paper proposes an automated preprocessing method designed to improve image classification accuracy when training with noisy data. The study reveals that while there is a relationship between rank and accuracy, this relationship

| World Heritage | | | | Correlation analysis before and after applying correlation | |
|---|---|---|---|---|---|
| threshold | number | Before | After | t-value | p-value |
| 0% | 5000 | 0.791 | 0.794 | -0.1464311571086384 | 0.8845606242583837 |
| 5% | 4521 | 0.878 | 0.891 | Mean value of prior learning model | |
| 10% | 4343 | 0.899 | 0.889 | number | accuracy |
| 15% | 4163 | 0.915 | 0.914 | 3500 | 0.785 |
| 20% | 4033 | 0.932 | 0.936 | | |
| 25% | 3899 | 0.946 | 0.952 | | |
| 30% | 3781 | 0.957 | 0.958 | | |
| 35% | 3661 | 0.955 | 0.957 | | |
| 40% | 3566 | 0.961 | 0.960 | | |
| 45% | 3431 | 0.970 | 0.976 | | |
| 50% | 3308 | 0.973 | 0.981 | | |
| 55% | 3159 | 0.979 | 0.980 | | |
| 60% | 3009 | 0.983 | 0.980 | | |
| 65% | 2771 | 0.990 | 0.990 | | |
| 70% | 2276 | 0.985 | 0.996 | | |
| 75% | 1425 | 0.985 | 0.988 | | |

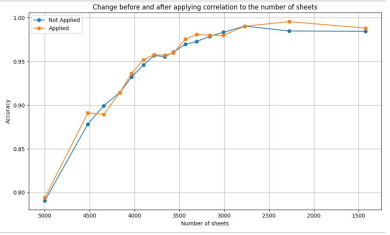


Fig. 5. Models with high accuracy Correlation analysis before and after applying correlation (Category:World Heritage)

| color | | | | Correlation analysis before and after applying correlation | |
|---|---|---|---|---|---|
| threshold | number | Before | After | t-value | p-value |
| 0% | 5000 | 0.319 | 0.305 | 0.01974101064009831 | 0.9843811565914916 |
| 5% | 4752 | 0.320 | 0.320 | Mean value of prior learning model | |
| 10% | 4515 | 0.321 | 0.344 | number | accuracy |
| 15% | 4229 | 0.358 | 0.365 | 3500 | 0.282 |
| 20% | 3935 | 0.364 | 0.388 | | |
| 25% | 3700 | 0.418 | 0.413 | | |
| 30% | 3429 | 0.416 | 0.434 | | |
| 35% | 3187 | 0.433 | 0.399 | | |
| 40% | 2943 | 0.473 | 0.454 | | |
| 45% | 2696 | 0.485 | 0.498 | | |
| 50% | 2454 | 0.502 | 0.516 | | |
| 55% | 2225 | 0.519 | 0.519 | | |
| 60% | 1993 | 0.505 | 0.520 | | |
| 65% | 1743 | 0.573 | 0.614 | | |
| 70% | 1539 | 0.629 | 0.571 | | |
| 75% | 1313 | 0.646 | 0.609 | | |

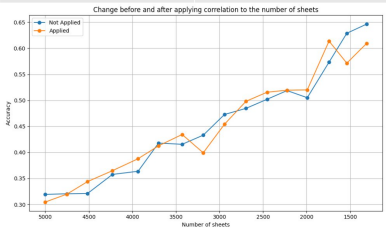


Fig. 6. Models with low accuracy Correlation analysis before and after applying correlation (Category:color)

does not significantly impact overall model performance. Additionally, excessively raising the threshold can improve apparent accuracy but may degrade the model's functionality. The research also shows that using multiple pre-trained models is effective, but the accuracy converges after a certain point.

## References

[1] G. Carneiro, A. B. Chan, and P. J. Moreno, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, no.7, pp.1204-1217, 2007.

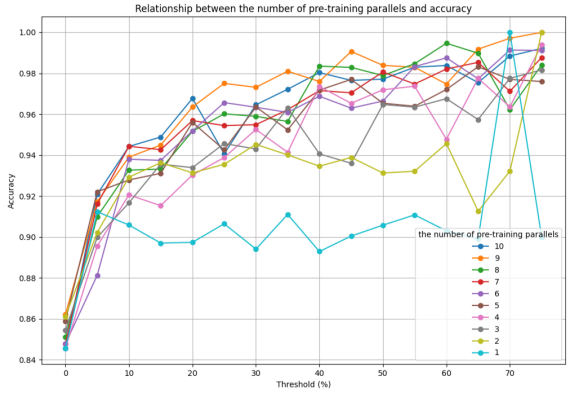


Fig. 7. Accuracy of pre-trained model per number of concurrences (Category:flower)

TABLE IV
ACCURACY OF PRE-LEARNING MODEL AND FINAL MODEL

| Model | Fashion | Art Style | Dog Breed | Car Type | Capital | Color | World Heritage | Geo Features | Hairstyle | Cuisine |
|---|---|---|---|---|---|---|---|---|---|---|
| Model1 | 0.403 | 0.319 | 0.774 | 0.489 | 0.388 | 0.256 | 0.774 | 0.319 | 0.489 | 0.388 |
| Model2 | 0.404 | 0.339 | 0.760 | 0.501 | 0.423 | 0.266 | 0.760 | 0.339 | 0.501 | 0.423 |
| Model3 | 0.390 | 0.334 | 0.771 | 0.513 | 0.395 | 0.270 | 0.780 | 0.359 | 0.489 | 0.388 |
| Model4 | 0.393 | 0.350 | 0.795 | 0.500 | 0.414 | 0.228 | 0.766 | 0.331 | 0.487 | 0.394 |
| Model5 | 0.367 | 0.309 | 0.763 | 0.477 | 0.377 | 0.225 | 0.784 | 0.343 | 0.486 | 0.372 |
| Model6 | 0.369 | 0.339 | 0.761 | 0.505 | 0.374 | 0.272 | 0.766 | 0.329 | 0.516 | 0.408 |
| Model7 | 0.391 | 0.331 | 0.767 | 0.488 | 0.408 | 0.265 | 0.782 | 0.341 | 0.468 | 0.370 |
| Model8 | 0.376 | 0.305 | 0.767 | 0.496 | 0.373 | 0.258 | 0.788 | 0.366 | 0.458 | 0.372 |
| Model9 | 0.368 | 0.325 | 0.781 | 0.488 | 0.352 | 0.245 | 0.771 | 0.339 | 0.475 | 0.413 |
| Model10 | 0.382 | 0.319 | 0.768 | 0.488 | 0.400 | 0.234 | 0.782 | 0.346 | 0.474 | 0.376 |
| Final | 0.464 | 0.421 | 0.893 | 0.573 | 0.456 | 0.302 | 0.904 | 0.398 | 0.561 | 0.436 |

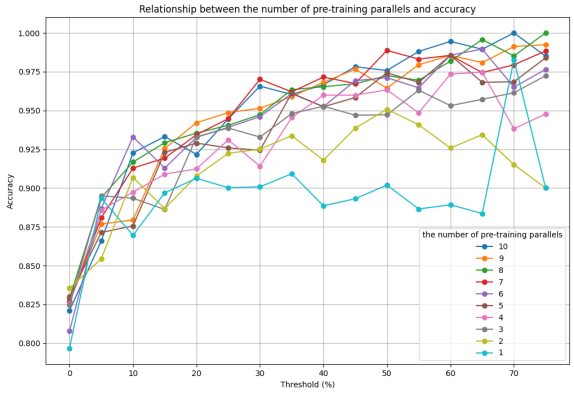


Fig. 8. Accuracy of pre-trained model per number of concurrences (Category:household furniture)

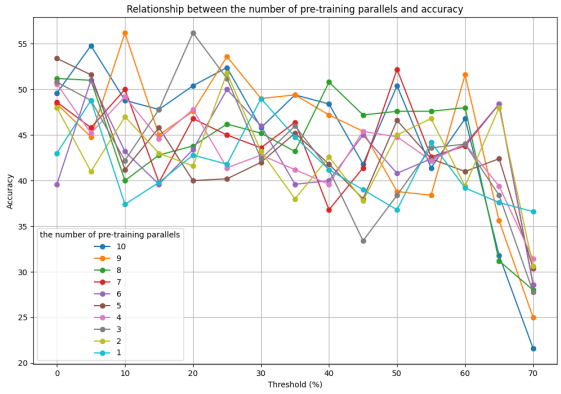


Fig. 10. Accuracy of pre-trained model accuracy calculated using the CIFER-100 dataset (Category:household furniture)

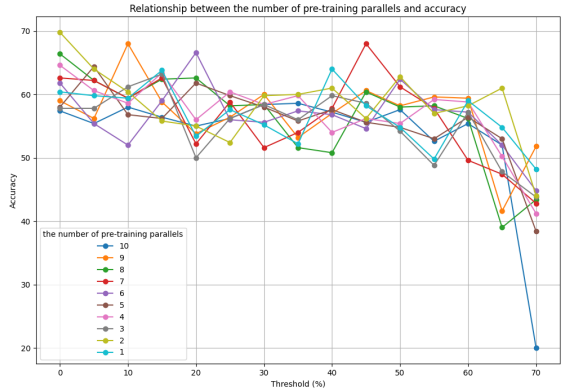


Fig. 9. Accuracy of pre-trained model accuracy calculated using the CIFER-100 dataset (Category:flowers)


[2] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training for semi-supervised image recognition," *Proceedings of the European Conference on Computer Vision (ECCV)*, pp.135-152, 2018.

[3] Y. Ouali, C. Hudelot, and M. Tami, "An overview of deep semi-supervised learning," *arXiv preprint arXiv:2006.05278*, 2020.

[4] D. Song, Y. Zhang, X. Shan, J. Cui, and H. Wu, "'Over-Learning' phenomenon of wavelet neural networks in remote sensing image classifications with different entropy error functions," *Entropy*, vol.19, no.10, pp.526, 2017.

[5] A. S. Potapov, "Principle of representational minimum description length in image analysis and pattern recognition," *Pattern Recognition and Image Analysis*, vol.22, no.1, pp.71-80, 2012.

[6] S. Bujwid, A. Pieropan, H. Azizpour, and A. Maki, "An analysis of over-sampling labeled data in semi-supervised learning with FixMatch," *arXiv preprint arXiv:2204.06851*, 2022.

[7] J. N. Eckardt, M. Bornhäuser, K. Wendt, et al., "Semi-supervised learning in cancer diagnostics," *Frontiers in Oncology*, vol.12, pp.763847, 2022.

[8] Y. Duan, Z. Zhao, L. Qi, L. Zhou, L. Wang, and Y. Shi, "Towards semi-supervised learning with non-random missing labels," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.16121-16131, 2023.

[9] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang, "Deep learning for detecting objects in remote sensing images," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2691-2699, 2015.

[10] N. Towghi and B. Javidi, "Image recognition in the presence of non-Gaussian noise with unknown statistics," *Journal of the Optical Society of America A (JOSA A)*, vol.18, no.8, pp.2007-2015, 2001.

[11] D. Sharma, R. Shukla, A. K. Giri, and others, "A brief review on search engine optimization," *Proceedings of the 9th International Conference on Communication Systems and Networks (COMSNETS)*, pp.320-326, 2019.

[12] V. K. Gunjan, M. Kumari, A. Kumar, and A. A. Rao, "Search engine optimization with Google," *Asian Journal of Engineering and Technology (AJET)*, vol.1, no.1, pp.22-30, 2012.

[13] S. Tollari, S. Glotin, and J. Le Maitre, "Enhancement of Textual Images Classification Using Segmented Visual Contents for Image Search Engine," *Multimedia Tools and Applications*, vol.25, no.3, pp.405–417, 2005.

[14] D. Müller, I. Soto-Rey, and F. Kramer, "An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks," *IEEE Access*, vol.10, pp.12345-12356, 2022.

[15] K. Wada, T. NoRo, "Consideration on the Influence of Weight Functions and the Scale for Robust Regression Estimator," in *Statistical Research Bulletin* , vol.78, pp.101-114, 2019.(in Japanese)