Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

Article - e007

Image-Based User Feature Classification on Social Media Using Machine Learning

Nguyen Thi Hoi[1], **Tran Thi Nhung**[2], **Dam Gia Manh**[3]

[1,2] Faculty of Information System and E- Commerce, Thuongmai University, Vietnam
[3]Department of Information Technology, Dong Do University, Vietnam

Received: 05/04/2026 **Revision Received**: 01/05/2026 **Accepted**: 08/05/2026

ABSTRACT

This study presents a robust machine learning framework for image-based user feature classification on social media, with a particular emphasis on applications in e-commerce environments. To address the critical challenge of limited labeled data within large-scale and unstructured image repositories, the proposed approach integrates semi-supervised learning techniques to enhance model generalization and data utilization efficiency. Three widely adopted machine learning algorithms, Decision Tree, Random Forest, and Support Vector Machine, were systematically evaluated using K-fold cross-validation and standard performance metrics, including accuracy, precision, recall, and F1-score. Experimental results demonstrate that the Random Forest model outperforms other approaches, achieving an accuracy of 0.9468 and an F1-score of 0.9456, while maintaining strong computational efficiency and robustness. These findings underscore the effectiveness of ensemble learning methods in handling high-dimensional, imbalanced datasets. The proposed framework offers significant practical implications for user profiling, personalized recommendation systems, and targeted marketing strategies in modern digital ecosystems.

KEYWORDS: Image Data, Feature Classification, Machine Learning, User Profile

1. INTRODUCTION

In the era of digital transformation and the rapid proliferation of multimedia data, image-based user feature classification on social media using machine learning has emerged as a critical research domain at the intersection of Computer Vision and Data Mining. Social media platforms such as Facebook, Instagram, and TikTok continuously generate massive volumes of visual content, offering valuable insights into user behavior, preferences, and consumption patterns. Recent advances in Deep Learning have significantly enhanced image classification performance, with architectures such as AlexNet and ResNet achieving remarkable breakthroughs in visual recognition and representation learning (Krizhevsky et al., 2012; He et al., 2016).

Despite these advances, traditional supervised learning approaches heavily rely on large-scale labeled datasets, which are costly, time-consuming, and often impractical to obtain in real-world social media contexts. To address this limitation, Semi-Supervised Learning (SSL) has emerged as an effective paradigm that leverages both labeled and unlabeled data to improve model performance and generalization (Chapelle et al., 2006; Zhu & Goldberg, 2009). Recent state-of-the-art methods, such as FixMatch and Mean Teacher, demonstrate that SSL can achieve competitive or even superior performance compared to fully supervised approaches under limited labeled data conditions (Sohn et al., 2020; Tarvainen & Valpola, 2017).

In Vietnam, the rapid expansion of social media usage has led to an explosion of user-generated multimedia content. According to DataReportal (2024), tens of millions of users actively create and share visual data on a daily basis. However, the lack of standardized labeled datasets and context-aware machine learning models remains a significant barrier to effectively exploiting this rich data source. From an application perspective, image-based user feature classification plays a pivotal role in E-commerce systems, particularly in user profiling, personalized recommendation, and targeted marketing. Major platforms such as Shopee and Lazada increasingly leverage multimodal data to enhance user engagement and conversion rates (Ricci et al., 2015; Jannach et al., 2016).

Furthermore, the dominance of Big Data and the prevalence of unstructured multimedia information have intensified the demand for scalable and efficient classification approaches. In this context, SSL provides a practical and cost-effective solution by exploiting large volumes of unlabeled data to improve learning efficiency and predictive performance (Goodfellow et al., 2016; Oliver et al., 2018).

Motivated by these challenges and opportunities, this study proposes a machine learning framework for image-based user feature classification on social media, with a focus on semi-

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

supervised learning techniques. The study aims to (i) develop an efficient and scalable classification framework, (ii) evaluate the effectiveness of SSL under limited labeled data conditions, and (iii) demonstrate its practical applicability in e-commerce and digital marketing contexts.

2. LITERATURE REVIEW

2.1 Data Classification and Social Media Images

Social media images represent a dominant form of user-generated content, widely shared on platforms such as Facebook, Instagram, and TikTok. Unlike curated datasets, these images are unstructured, heterogeneous, and context-rich, capturing not only visual elements but also user intentions, social signals, and cultural contexts. As a result, they provide valuable insights into user identity and behavior (Highfield & Leaver, 2016; Gelli et al., 2015).

However, social media images present significant challenges, including noise, redundancy, high variability, and semantic ambiguity. Moreover, the scarcity of labeled data limits the effectiveness of traditional supervised learning approaches. To address these issues, recent studies increasingly adopt SSL, which leverages both labeled and unlabeled data to improve model generalization (Chapelle et al., 2006; Zhu & Goldberg, 2009).

Image classification in this context involves assigning semantic labels based on visual content, context, or user behavior. Common approaches include content-based, semantic, affective, and context-aware classification. Advances in deep learning, particularly CNNs such as AlexNet and ResNet, have significantly improved performance (Krizhevsky et al., 2012; He et al., 2016).

More recently, multimodal models like CLIP integrate visual and textual data to enhance classification. Nevertheless, challenges such as noisy labels and data imbalance persist, making SSL a promising solution for scalable and robust applications.

2.2. Machine Learning Techniques for Image Classification and E-commerce Applications

Image classification techniques have evolved significantly, from traditional feature-engineering approaches to advanced deep learning paradigms. Early methods relied on handcrafted features such as SIFT and HOG, combined with classifiers like SVM and KNN. While effective in controlled settings, these approaches were limited in scalability and representation for complex, large-scale datasets.

The emergence of deep learning, particularly CNNs, marked a major breakthrough. Architectures such as AlexNet, VGG, and ResNet automatically learn hierarchical features from raw images, significantly improving classification accuracy and generalization (Krizhevsky et al., 2012; Simonyan & Zisserman, 2015; He et al., 2016).

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

More recently, Vision Transformers (ViT) have introduced a new paradigm by leveraging self-attention to capture global image dependencies (Dosovitskiy et al., 2021). At the same time, multimodal learning approaches integrate visual data with textual information such as captions and hashtags. Vision-language models like CLIP demonstrate strong generalization and adaptability in open-domain environments (Radford et al., 2021).

Modern frameworks also widely adopt transfer learning and fine-tuning, utilizing pre-trained models on datasets such as ImageNet to improve efficiency and reduce computational cost (Goodfellow et al., 2016). Additionally, self-supervised and SSL methods enable effective use of large-scale unlabeled data, which is particularly valuable in social media contexts (Chen et al., 2020; Sohn et al., 2020).

From an application perspective, image-based classification is essential in E-commerce platforms such as Shopee and Amazon, supporting user profiling, personalized recommendation, and marketing optimization. Integrating visual features into recommender systems has been shown to improve prediction accuracy and user satisfaction, especially when combined with multimodal data sources (Jannach et al., 2016; Wu et al., 2022).

2.3. Semi-Supervised Learning Techniques for Image Classification

Semi-Supervised Learning has emerged as an effective paradigm for image classification, particularly in social media environments where labeled data are limited but unlabeled data are abundant. By leveraging both data types, SSL improves model generalization and reduces reliance on costly annotation processes (Chapelle et al., 2006; Zhu & Goldberg, 2009). This makes it highly suitable for image-based user feature classification tasks involving large-scale and noisy data.

Traditional SSL approaches include self-training, co-training, and consistency regularization. Self-training iteratively assigns pseudo-labels to unlabeled data, while co-training utilizes multiple feature views to enhance learning. Consistency regularization enforces stable predictions under data perturbations, improving robustness (Oliver et al., 2018). However, these methods may be limited in complex, high-dimensional image tasks.

Recent advances in deep learning have significantly enhanced SSL performance. Teacher–student frameworks such as Mean Teacher improve prediction stability (Tarvainen & Valpola, 2017), while FixMatch combines pseudo-labeling with strong augmentation to achieve high performance under limited labeled data (Sohn et al., 2020). Additionally, self-supervised approaches like SimCLR and MoCo learn robust feature representations without explicit labels (Chen et al., 2020; He et al., 2020).

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

More recent developments extend SSL to multimodal and large-scale settings. Models like CLIP align visual and textual representations (Radford et al., 2021), while Noisy Student Training enhances scalability (Xie et al., 2020). Overall, SSL provides a scalable and robust solution for image-based classification in social media and E-commerce applications.

3. RESEARCH MODEL PROPOSAL

3.1. Problem Statement

With the rapid growth of E-commerce and the widespread use of social media platforms such as Facebook, Instagram, and TikTok, massive volumes of user-generated images are continuously produced. These images contain valuable information about user behavior, preferences, and personal characteristics, which are essential for recommendation systems and personalized services. However, this task faces several challenges, including unstructured and noisy data, lack of labeled samples, and imbalanced long-tail distributions.

Formally, given a dataset $D = \{(x_i, y_i)\}_{i=1}^{n_l} \cup \{x_j\}_{j=1}^{n_u}$, where labeled data are limited (n_l) and (n_u)unlabeled data dominate ($n_u \gg n_l$), the objective is to learn a model (f(x)) that accurately predicts user labels while effectively leveraging unlabeled data.

The output labels may include user interests, customer segments, and consumption behaviors, supporting applications such as personalized recommendation and targeted marketing on platforms like Shopee and Amazon. Therefore, integrating deep learning with Semi-Supervised Learning is a promising approach to address these challenges and improve classification performance.

3.2. The proposal research framework

This study proposes a comprehensive machine learning framework for image-based user feature classification on social media using machine learning, with a particular focus on scalability and applicability in E-commerce environments. The framework is designed as a structured pipeline consisting of seven sequential stages, enabling the effective exploitation of large-scale and unstructured image data.

The proposed workflow includes: (1) Data Acquisition → (2) Data Preprocessing → (3) Data Labeling → (4) Data Splitting → (5) Model Training → (6) Model Optimization → (7) Model Evaluation.

Evaluation Protocol and Model Assessment: Model performance is evaluated using a confusion matrix (TP, TN, FP, FN) to analyze classification outcomes and error distribution. Based on this, key metrics including accuracy, precision, recall, and F1-score are applied to provide a comprehensive assessment, particularly for imbalanced social media data. In addition,

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

computational efficiency is measured through training time to assess model scalability and practicality. Comparative experiments across multiple algorithms evaluate both predictive performance and computational cost.

Overall, this framework ensures a balanced evaluation of effectiveness and efficiency, supporting real-world applications such as user profiling, personalized recommendation, and targeted marketing in digital environments.

4. RESULTS AND ANALYSIS

4.1. Construction of the Experimental Dataset

The experimental dataset was constructed to support image-based user feature classification on social media using machine learning, with particular consideration of semi-supervised learning scenarios. Data were collected from 500 Facebook user accounts, with each user contributing approximately 20 public images, resulting in an initial dataset of 10,000 images. After removing duplicate, low-quality, and corrupted images, the final dataset consisted of 9,748 valid images for analysis.

To ensure consistency, all images were resized to 128 × 128 pixels, converted to the RGB color space, and stored in JPEG format. Each image was then transformed into a pixel matrix and further converted into a flattened feature vector. Pixel values were normalized to the range [0, 1] by dividing by 255, which improves computational stability and supports efficient model training.

For data organization, each user was assigned a unique identifier from 1 to 500. Images were stored in user-specific directories, and file naming followed the format "a_b", where *a* represents the user ID and *b* denotes the image index. The processed data were then saved in Data.csv, which contains the complete dataset, including image identifiers, filenames, and normalized pixel values.

The dataset was annotated into 52 semantic categories, including representative classes such as beach, fashion, food, selfie, and sport. A hybrid annotation strategy was applied to improve both labeling quality and scalability. Manual labeling was conducted based on visual content and contextual interpretation, while automated labeling was supported by machine learning procedures evaluated through K-fold cross-validation. To assess label consistency, a binary flag was introduced: samples with matching manual and automated labels were assigned a value of 1, whereas mismatched samples were assigned 0.

The experimental environment included an 11th Gen Intel Core i7-1185G7 processor, 32 GB RAM, and the Windows 11 operating system. The implementation was developed in Python using Visual Studio Code 2020. The experimental workspace included *Data.csv* and three algorithm-specific scripts: *Decision_Tree.py*, *Random_Forest.py*, and *SVM.py*. Several Python libraries were

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

used, including NumPy for numerical computation, Pandas for data manipulation and CSV processing, Scikit-learn for model implementation and evaluation, and Matplotlib and Seaborn for visualization and performance analysis.

Overall, the structured preprocessing, labeling, and experimental setup provide a reliable and reproducible foundation for evaluating image-based user classification models in social media and E-commerce contexts.

Table1: Sample Data

User ID	Image Name	Category Label	Pixel Vector (Flattened)	Consistency Flag
U001	1_01	selfie	[0.12, 0.45, …]	1
U001	1_02	food	[0.67, 0.23, …]	1
U002	2_01	beach	[0.34, 0.78, …]	0
U002	2_02	fashion	[0.56, 0.12, …]	1
…	…	…	…	…

4.2. Dataset Statistics and Imbalance Analysis

To better understand the characteristics of the dataset, statistical analysis was conducted on the distribution of images across the 52 semantic categories. The dataset consists of 9,748 images collected from 500 users, with an average of approximately 19.5 images per user. However, the distribution across classes is highly imbalanced, which is a common issue in social media datasets.

Specifically, certain categories such as selfie, food, photo, and fashion dominate the dataset due to their high frequency in user-generated content, while others such as passport, construction, and tapeworm are underrepresented. This results in a long-tail distribution, where a small number of classes contain the majority of samples, while many classes have very few instances.

This imbalance can negatively impact model performance, leading to biased predictions toward majority classes and poor generalization on minority classes. To address this issue, several strategies are considered, including data augmentation, class weighting, and resampling techniques. Additionally, evaluation metrics such as F1-score and macro-averaging are prioritized to ensure fair performance assessment across all classes.

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

Overall, the dataset reflects real-world social media characteristics, where imbalance and noise are inherent, thus providing a realistic benchmark for evaluating machine learning models in practical applications.

4.3 Experimental Scenario

Step 1: Data Collection: A total of 612 Facebook user accounts were collected and filtered to 500 valid users. Public images were automatically extracted and organized into user-specific folders in chronological order.

Step 2: Data Preprocessing: Images were resized to 128×128, converted to RGB, and transformed into flattened vectors with normalized pixel values in [0,1]. Data were stored in CSV format.

Step 3: Labeling and Exploration: Images were annotated into 52 categories using Roboflow. The dataset was analyzed and split into 80% training and 20% testing, with K-fold cross-validation applied.

Step 4: Data Splitting: The training set was divided into 5 folds (4 for training, 1 for validation), resulting in 7,798 training, 1,950 validation, and 1,950 test samples.

Step 5: Model Training: Models including Decision Tree, Random Forest, and SVM were trained and tuned using validation data.

Step 6: Evaluation: Models were evaluated using accuracy, precision, recall, and F1-score. The best model was selected for real-world applications.

4.4. Results and Discussion

The experimental results provide comprehensive insights into the performance, robustness, and computational efficiency of the evaluated machine learning models for image-based user feature classification on social media.

Cross-Validation Performance and Model Stability

The K-fold cross-validation results reveal that the Random Forest model consistently achieves the highest classification accuracy across all folds, ranging from 0.9395 to 0.9542. The relatively narrow variation across folds indicates strong stability and robustness, suggesting that the model is not sensitive to variations in training data distribution. The average cross-validation accuracy of Random Forest (0.9480) significantly outperforms both Decision Tree and SVM, highlighting the effectiveness of ensemble learning in reducing variance and improving generalization.

In contrast, the SVM model demonstrates moderately strong performance, with an average accuracy of approximately 0.9018 across folds. Although its performance is relatively stable, it remains consistently lower than Random Forest. Meanwhile, the Decision Tree model exhibits the

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

weakest performance (approximately 0.79), reflecting its susceptibility to overfitting and limited capability in handling high-dimensional and noisy image data.

Generalization Performance on Test Data

The test results further confirm the generalization capability of the models. Random Forest maintains a high accuracy of 0.9468 on the test set, showing only a marginal decrease compared to cross-validation results. This minimal performance gap indicates that the model effectively avoids overfitting and generalizes well to unseen data.

Similarly, the SVM model achieves a test accuracy of 0.8942, which is consistent with its cross-validation performance, suggesting stable generalization behavior. However, its overall predictive power remains lower than that of Random Forest. The Decision Tree model records a test accuracy of 0.7997, confirming its limited generalization ability and reinforcing the need for more robust learning strategies in complex image classification tasks.

Comparative Analysis of Evaluation Metrics

A detailed comparison of evaluation metrics demonstrates the clear superiority of Random Forest across all performance indicators. Specifically, Random Forest achieves an accuracy of 0.9468, precision of 0.9498, recall of 0.9402, and an F1-score of 0.9456. These results indicate not only high overall accuracy but also a balanced trade-off between false positives and false negatives, which is particularly important in imbalanced datasets.

The SVM model achieves relatively balanced performance (F1-score = 0.8954), indicating its effectiveness as a baseline classifier. However, its performance remains inferior to Random Forest, particularly in terms of recall, suggesting potential limitations in capturing complex feature distributions. The Decision Tree model, although computationally simple, shows lower precision and recall values, confirming its limitations in high-dimensional feature spaces.

Computational Efficiency and Scalability

In addition to predictive performance, computational efficiency is a critical factor for real-world deployment. The results indicate that Random Forest achieves the fastest training time (7.19 seconds), significantly outperforming both Decision Tree (123.74 seconds) and SVM (170.56 seconds).

This finding is particularly noteworthy, as Random Forest simultaneously delivers the highest predictive performance and the lowest computational cost. In contrast, SVM, despite achieving reasonable accuracy, requires substantially longer training time, making it less suitable for large-scale or real-time applications. The Decision Tree model, while faster than SVM, still fails to achieve competitive performance.

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

The superior balance between accuracy and training efficiency makes Random Forest highly suitable for scalable deployment in E-commerce systems, where large volumes of social media data must be processed efficiently.

Table 1. Comparison among three algorithms on each fold and tests

Fold	Decision Tree	Random Forest	SVM
Fold 1	0.7905	0.9531	0.9016
Fold 2	0.7884	0.9460	0.9002
Fold 3	0.7932	0.9542	0.9051
Fold 4	0.7891	0.9395	0.8998
Fold 5	0.7915	0.9473	0.9021
AVG	**0.7905**	**0.9480**	**0.9018**
Test	0.7997	0.9468	0.8942

(Source: Results by python)

Table 2. The Results of three algorithms

Algorithms	Accuracy on folds	Accuracy on test	Precision	Recall	F1	Time Train
Decision Tree	0.7905	0.7997	0.8012	0.7793	0.7915	123.74s
Random Forest	0.9480	0.9468	0.9498	0.9402	0.9456	7.19s
SVM	0.9018	0.8942	0.9009	0.8905	0.8954	170.56s

(Source: Results by python)

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

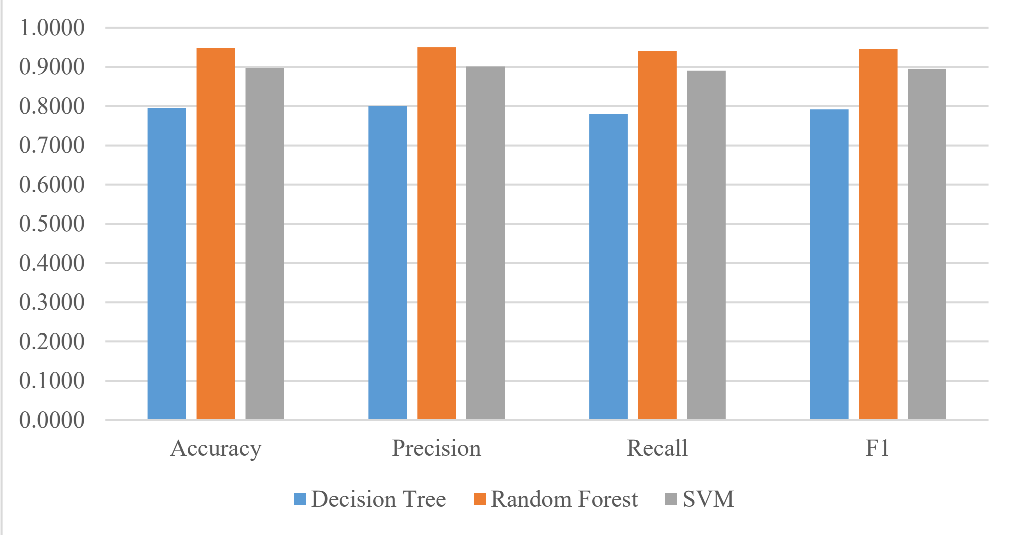

Fig 1. The Results of three algorithms

4.5 Discussion and Implications for E-commerce Applications

From an application perspective, the experimental findings highlight the importance of robustness, scalability, and computational efficiency in real-world digital ecosystems. On large-scale E-commerce platforms such as Shopee and Amazon, image-based user feature classification plays a key role in user profiling, personalized recommendation, and targeted marketing. These applications require models that achieve high accuracy while maintaining efficient performance in dynamic environments.

The superior performance of Random Forest demonstrates its strong ability to handle high-dimensional, noisy, and imbalanced social media image data. Its ensemble structure reduces variance and enhances robustness, while its relatively low training time supports rapid model updating for real-time systems.

Compared to single models such as Decision Tree, Random Forest significantly improves predictive accuracy and generalization. Although SVM shows stable performance, its higher computational cost limits scalability. In contrast, Random Forest provides a more effective balance between accuracy and efficiency, making it suitable for large-scale deployment.

Bias–variance analysis further confirms these findings. Decision Tree tends to overfit, exhibiting high variance and unstable performance. SVM achieves moderate generalization but remains computationally expensive. Random Forest, however, achieves the best balance, reducing variance while maintaining low bias and consistent performance.

Practically, integrating Random Forest into recommendation systems can enhance user segmentation, improve personalization, and support data-driven marketing strategies. Overall, this

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

study confirms that Random Forest is a robust, scalable, and efficient solution for image-based user classification in modern E-commerce environments.

5. CONCLUSION

This study proposes is leveraging Semi-Supervised Learning to address limited labeled data. The results show that Random Forest outperforms other models in terms of accuracy, F1-score, and computational efficiency, demonstrating strong robustness and generalization.

The findings highlight the importance of handling high-dimensional, noisy, and imbalanced image data in real-world environments. The integration of preprocessing, hybrid labeling, and optimization techniques significantly enhances model performance.

From both theoretical and practical perspectives, this study confirms the effectiveness of ensemble learning for scalable applications in social media analytics and E-commerce systems, particularly in user profiling and personalized recommendation.

ACKNOWLEDGMENTS

The authors declare that no financial or institutional support was received for this research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data used in this study are not publicly available due to privacy and platform restrictions but may be available from the corresponding author upon reasonable request.

REFERENCES:

1. Bakhshi, S., Shamma, D. A., & Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on Instagram. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 965–974). https://doi.org/10.1145/2556288.2557403
2. Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443.
3. Chapelle, O., Schölkopf, B., & Zien, A. (2006). Semi-supervised learning. MIT Press.

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

4. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. Proceedings of the 37th International Conference on Machine Learning (ICML).
5. DataReportal. (2024). Digital 2024: Vietnam. https://datareportal.com
6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR).
7. Gelli, F., Uricchio, T., Bertini, M., Del Bimbo, A., & Chang, S.-F. (2015). Image popularity prediction in social media using sentiment and context features. Proceedings of the ACM International Conference on Multimedia.
8. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.
10. Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media. Communication Research and Practice, 2(1), 47–62. https://doi.org/10.1080/22041451.2016.1155332
11. Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. In Proceedings of the International AAAI Conference on Web and Social Media (ICWSM).
12. Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2016). Recommender systems: An introduction. Cambridge University Press.
13. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.
14. Oliver, A., Odena, A., Raffel, C., Cubuk, E. D., & Goodfellow, I. (2018). Realistic evaluation of deep semi-supervised learning algorithms. Advances in Neural Information Processing Systems, 31.
15. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning (ICML).
16. Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems handbook (2nd ed.). Springer.

ISSN: 3139-3527 | April-June-Issue, Vol. 1, No. 2 (2026) | DOI: 10.66261/02rp8324

17. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. ICLR.
18. Sohn, K., Berthelot, D., Li, C.-L., Zhang, Z., Carlini, N., Cubuk, E. D., Kurakin, A., & Raffel, C. (2020). FixMatch: Simplifying semi-supervised learning with consistency and confidence. NeurIPS.
19. Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In Advances in Neural Information Processing Systems (NeurIPS).
20. Wu, L., Sun, P., Hong, R., Wang, Y., & Wang, M. (2022). Multimodal recommendation systems: A survey. ACM Transactions on Information Systems, 40(2), 1–38.
21. Xie, Q., Luong, M. T., Hovy, E., & Le, Q. V. (2020). Self-training with noisy student improves ImageNet classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
22. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2020). Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys, 52(1), 1–38.
23. Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1–130.