VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

OVERFITTING MITIGATION AND TRAINING OPTIMIZATION STRATEGIES IN DEEP LEARNING-BASED IMAGE CLASSIFICATION: A SYSTEMATIC REVIEW

Budiman[1,2*]; Arief Setyanto[1]; Andi Sunyoto[1]; Dhani Ariatmanto[1]

Informatics[1]
Universitas Amikom Yogyakarta, Yogyakarta, Indonesia[1]
https://home.amikom.ac.id/[1]
budiman@students.amikom.ac.id[*]

Informatics[2]
Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia[2]
https://unibi.ac.id/[2]
budiman@unibi.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Overfitting remains a critical challenge in developing deep learning–based image classification models, particularly as modern architectures become increasingly complex and parameter-intensive. Although convolutional and transformer-based models have demonstrated strong predictive performance, their generalization capability depends strongly on the availability of large, well-annotated, and diverse training datasets. This condition is often difficult to achieve in real-world domains such as medicine, agriculture, and industrial inspection. Previous survey studies have examined various techniques related to image classification, data augmentation, and optimization strategies; however, these studies typically analyse individual approaches in isolation. As a result, opportunities remain to further synthesise the relationships among the underlying causes of overfitting, mitigation strategies, and training parameter configurations within a unified analytical perspective. This study employed the PRISMA framework to conduct a Systematic Literature Review (SLR). A total of 174 primary studies published between 2020 and 2025 were traced to address the gap. The review identifies three major sources of overfitting in image classification tasks: limited labeled data, model architecture complexity, and data and label quality issues. Based on these findings, the study synthesises the corresponding mitigation strategies reported in the literature. The main contribution of this study is not the proposal of a new theoretical taxonomy, but the systematic organisation and synthesis of methodological evidence into an integrated analytical framework. The resulting analytical framework relates the underlying causes of overfitting, mitigation strategies, training optimisation mechanisms, and parameter configuration practices within a unified perspective, thereby facilitating a more integrated interpretation of how these complementary aspects contribute to model generalisation in deep learning–based image classification.*

Keywords: *Deep Learning Optimization, Image Classification, Overfitting Mitigation, Systematic Literature Review, Training Stability.*

Intisari— *Overfitting tetap menjadi tantangan kritis dalam pengembangan model klasifikasi gambar berbasis deep learning, terutama seiring arsitektur modern yang semakin kompleks dan sarat parameter. Meskipun model berbasis konvolusional dan transformer telah menunjukkan kinerja prediktif yang kuat, kemampuan generalisasinya sangat bergantung pada ketersediaan dataset pelatihan yang besar, teranotasi dengan baik, dan beragam. Kondisi ini sering kali sulit dipenuhi dalam bidang-bidang dunia nyata seperti*

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

kedokteran, pertanian, dan inspeksi industri. Studi survei sebelumnya telah meneliti berbagai teknik terkait klasifikasi gambar, augmentasi data, dan strategi optimasi; namun, studi-studi tersebut umumnya menganalisis pendekatan-pendekatan tersebut secara terpisah. Akibatnya, sintesis komprehensif yang secara eksplisit mengaitkan penyebab mendasar overfitting dengan strategi mitigasi dan konfigurasi parameter pelatihan masih relatif terbatas. Penelitian ini menggunakan kerangka kerja PRISMA untuk melakukan Tinjauan Literatur Sistematis (SLR). Sebanyak 174 studi primer yang diterbitkan antara tahun 2020 dan 2025 ditelusuri untuk mengatasi kesenjangan tersebut. Tinjauan ini mengidentifikasi tiga sumber utama overfitting dalam tugas klasifikasi gambar: keterbatasan data berlabel, kompleksitas arsitektur model, serta masalah kualitas data dan label. Berdasarkan temuan ini, penelitian ini mensintesis strategi mitigasi yang sesuai sebagaimana dilaporkan dalam literatur. Kontribusi utama penelitian ini bukanlah usulan taksonomi teoretis baru, melainkan pengorganisasian dan sintesis sistematis bukti metodologis ke dalam kerangka kerja analitis yang terintegrasi. Kerangka kerja analitis yang dihasilkan mengaitkan penyebab mendasar dari overfitting, strategi mitigasi, mekanisme optimasi pelatihan, dan praktik konfigurasi parameter dalam satu perspektif terpadu, sehingga memfasilitasi interpretasi yang lebih terintegrasi mengenai bagaimana aspek-aspek yang saling melengkapi ini berkontribusi terhadap generalisasi model dalam klasifikasi gambar berbasis deep learning.

Kata Kunci: *Optimasi Deep Learning, Klasifikasi Citra, Mitigasi Overfitting, Tinjauan Pustaka Sistematis, Stabilitas Pelatihan.*

INTRODUCTION

In the development of deep learning-based image classification models, overfitting remains a major challenge because models tend to learn specific patterns—including noise—in the training data, thereby reducing their generalization ability [1], [2], [3], [4]. This problem becomes even more critical when labeled data is limited or of low quality, particularly in the fields of healthcare [5], [6], [7], [8], agriculture [1], [9], and industrial applications [10], [11]. Contributing factors include data limitations [1], [9], [12], architectural complexity [6], [10], [13], as well as annotation errors and variations in image quality [14], [15]. Meanwhile, knowledge and feature distillation techniques have become widely used [16], [17], [18], [19], [20], but their contribution to mitigating overfitting has not yet been systematically and comprehensively analyzed.

Various strategies have been developed, including data augmentation [21], [22], [23], transfer learning [24], [25], architectural regularization [26], [27], [28], and adaptive optimization and schedulers [29], [30]. However, previous studies generally discuss these approaches separately or in the context of specific applications, so the relationship between the causes of overfitting, mitigation strategies, optimization mechanisms, and training parameter configurations remains scattered across various publications. Therefore, this SLR develops an analytical taxonomy that integrates data limitations [1], [12], architectural complexity [10], [31], annotation quality [14], [15], and the roles of knowledge and feature distillation [18], [32]. This framework is expected to help map the interrelationships among approaches and serve as a foundation for designing training processes that are more stable, efficient, and adaptive to data variations.

This study is designed to answer the following three main questions:

a. RQ1: How are methods used in the literature to overcome overfitting in image classification tasks?
b. RQ2: How can methods improve convergence during the image classification model training process?
c. RQ3: How can parameters reduce overfitting while maintaining or improving the performance of image classification models?

This study makes several important contributions. Building upon the analysis of previous survey studies discussed in the State of the Art section, this study systematically maps the recent literature. It examines techniques for mitigating overfitting, training strategies, and parameter configurations in the context of deep learning–based image classification. Second, this study proposes an analytical taxonomy that organises the literature around three major sources of overfitting: limited labeled data, model architecture complexity, and data or label quality. The taxonomy further maps the mitigation strategies reported in the literature, including data augmentation, transfer learning, architectural regularisation, optimization approaches, and knowledge or feature distillation techniques.

Third, the study identifies recent methodological trends reported in the literature, including advanced data augmentation approaches, architectural regularisation techniques, knowledge and feature distillation strategies, and increasingly adopted adaptive optimization methods. The analysis further explains how these strategies address different sources of overfitting and contribute to improved predictive robustness and training stability. In addition, this study synthesises the relationship between causal factors, mitigation techniques, convergence improvement mechanisms, and training parameter configurations, thereby providing practical insights for researchers and practitioners in designing more effective training pipelines. Finally, based on the comparative analysis of previous studies, this review highlights several research gaps that remain insufficiently addressed, including the need for more representative datasets, improved annotation quality, the development of more efficient model architectures, and greater consistency in evaluation standards across different application domains.

State of the Art

Advances in deep learning have improved the performance of image classification and various computer vision tasks. However, overfitting remains a challenge, particularly when training data is limited, imbalanced or noisy. A number of review studies have discussed mitigation strategies through specific approaches, such as low-shot learning, data augmentation, and the use of Generative Adversarial Networks (GANs). Study [14], for example, examines low-shot object detection and highlights the role of transfer learning, meta-learning, and prototype-based modelling in conditions of limited labelled data. Meanwhile, [22] focuses on data augmentation techniques for classification and segmentation, whilst [33] discusses the use of GANs to generate synthetic images for the analysis of skin lesions. These three studies make important contributions within their respective contexts, but their scope remains specific, whether in terms of task type, mitigation techniques, or application domains.

Based on the studies analysed, comprehensive research into mitigating overfitting in deep learning-based image classification remains relatively limited. The available literature tends to evaluate the effectiveness of a technique in isolation, without systematically linking the factors causing overfitting, mitigation strategies, optimisation mechanisms, and training parameter configurations. In fact, these aspects are interrelated in determining generalisation ability, the stability of the training process, and model performance. These limitations mean that understanding of overfitting mitigation remains scattered across various approaches and research contexts, as summarised in Table 1.

Based on the comparison in Table 1, the focus of this study lies in an integrated literature synthesis approach, rather than in proposing specific mitigation techniques. This study develops a taxonomy that links the causes of overfitting, mitigation strategies, training optimisation, and parameter configuration within the context of deep learning-based image classification. Consequently, the study not only maps the techniques used to reduce overfitting but also provides an overview of the interrelationships between causal factors, mitigation mechanisms, and optimisation aspects that influence model generalisation and stability.

Table 1. Representative Review Studies Related to Overfitting Mitigation in Image Classification

Study	Focus	Scope	Limitation
[14]	Low-shot object detection	Detection tasks	Limited analysis of classification pipelines
[22]	Data augmentation	Classification & segmentation	Focus only on augmentation techniques
[33]	GAN augmentation	Medical imaging	Limited discussion on optimization and training parameters
This study	Overfitting mitigation taxonomy	Image classification	Integrates causes, mitigation strategies, and training optimization

Source: (Research Results, 2025)

MATERIALS AND METHODS

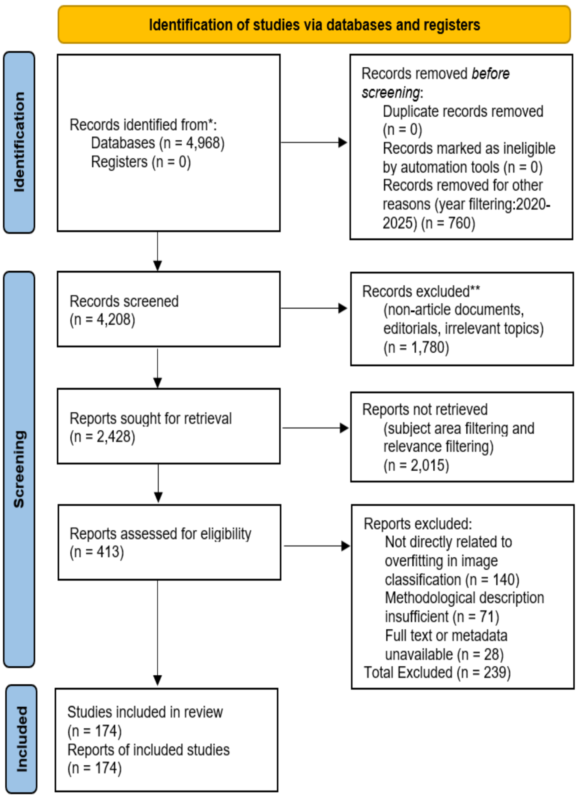
The Scopus database was used in this study for a Systematic Literature Review (SLR). The search was conducted on November 30, 2025, using the keywords “overfitting” and “image classification” via the query TITLE-ABS-KEY (“overfitting”) AND TITLE-ABS-KEY (“image classification”). Scopus was selected based on its multidisciplinary coverage, standardized metadata, and search capabilities that support the systematic review process. A focused search strategy was employed to ensure the relevance of the articles to the research objectives, although this approach may have excluded studies using other terms, such as image recognition, visual classification, model generalization, or regularization.

The article selection process was conducted in stages following the PRISMA framework, including the identification, screening, eligibility, and inclusion stages, as shown in Figure 1. Articles identified in the initial search were selected based on their titles and abstracts, then reviewed in full

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

text to ensure their alignment with the study's objectives and criteria. The use of Scopus as the sole literature source constitutes a methodological limitation, as publications indexed only in IEEE Xplore, the ACM Digital Library, or Web of Science may potentially go unidentified. Therefore, the results of this synthesis should be understood within the scope of the Scopus literature. Future research is advised to expand the range of database sources and use a more diverse combination of keywords and synonyms to reduce the potential for selection bias.



Source: (Research Results, 2025)

Figure 1. PRISMA 2020 flow diagram

A. Identification

The identification stage aimed to compile an initial corpus of literature on overfitting in image classification. The search conducted in the Scopus database produced 4,968 records across various research fields and publication types. To ensure that the literature reflected recent developments in deep learning, the publication period was limited to 2020–2025. This filtering removed 760 records published outside the specified range, resulting in 4,208 records that proceeded to the screening stage. This step ensured that the initial dataset represented current advances in deep learning architectures, including Convolutional Neural Networks (CNNs), EfficientNet, and Vision Transformers (ViTs).

B. Screening

The screening stage was performed to refine the dataset using basic quality and relevance criteria. At this stage, titles and abstracts were examined to exclude non-scientific documents such as editorials, technical reports, and publications that had not undergone peer review. A total of 1,780 records were removed during this process. After this filtering, 2,428 articles remained and were considered suitable for further evaluation. These studies were considered scientifically valid publications on deep learning–based image classification.

C. Eligibility Assessment

The remaining studies were evaluated based on their methodological relevance and alignment with the research objectives through screening in the fields of Computer Science, Engineering, and Mathematics, resulting in 413 articles. Next, a full-text review was conducted based on the established inclusion and exclusion criteria; articles that did not address overfitting in image classification, did not provide adequate methodological information, or had incomplete full text or metadata were excluded, leaving 174 articles.

This study did not use formal instruments such as CASP, JBI, or MMAT because the focus of the analysis was a synthesis of methodological developments, not an evaluation of intervention effectiveness or effect size calculations. The contribution of each study was assessed based on methodological relevance, completeness of reporting, and technical information that could be extracted for the coding and comparative synthesis processes. The screening, data extraction, and coding processes were conducted by the first author based on the established criteria and then reviewed with the supervisory team; differences in interpretation were resolved through discussion until a consensus was reached based on the study's objectives, selection criteria, and the evidence reported in the source articles.

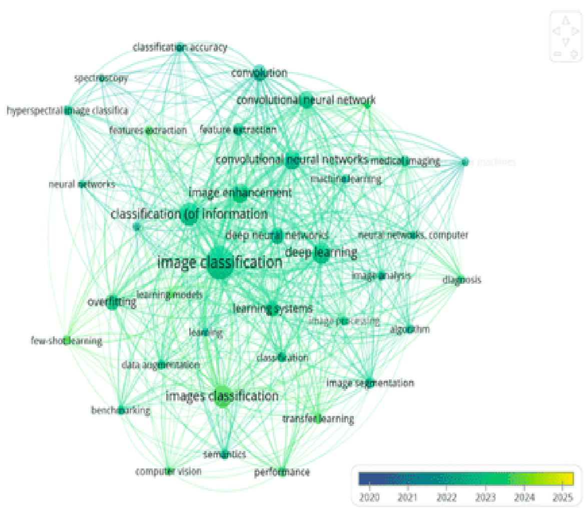
D. Final Inclusion

After all selection stages were completed, a total of 174 articles met the inclusion criteria and were designated as the final corpus for the Systematic Literature Review (SLR). All articles were analyzed to address RQ1, RQ2, and RQ3 through methodological coding that covered study characteristics, causes and mitigation strategies for overfitting, training optimization mechanisms,

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

parameter configurations, validation, and evaluation metrics. Data were extracted using a structured matrix containing bibliographic information, application domains, datasets, model architectures, mitigation methods, optimization, parameters, validation, and contributions to each research question. Of the 4,968 articles identified in the initial screening, the number was reduced to 4,208 after filtering by publication year, to 2,428 after reviewing titles and abstracts, to 413 after evaluating full texts, and finally to 174 articles based on all selection criteria in accordance with the PRISMA framework. Due to heterogeneity in the datasets, architectures, metrics, and validation procedures, the synthesis was conducted through descriptive and qualitative comparative analysis rather than quantitative meta-analysis.

The distribution of publications shows the dominance of Elsevier B.V. and IEEE, with approximately 30 and 26 articles, respectively, while Springer, Springer Science and Business Media Deutschland GmbH, and MDPI contributed approximately 9–13 articles. The bibliometric map in Figure 2 shows that research is centered on image classification, deep learning, and convolutional neural networks, which exhibit the strongest connectivity. These topics are related to feature extraction, image enhancement, classification accuracy, and generalization issues such as overfitting, data augmentation, and few-shot learning. Research developments are also moving toward transfer learning, image segmentation, and applications in the medical domain as well as more complex images, indicating a close relationship between architectural development, optimization strategies, and the need to improve model performance.



(Research Results, 2025)

Figure 2. Bibliometric Map

RESULTS AND DISCUSSION

A. RESULTS

1. Descriptive Analysis

An analysis of the frequency of methodological keywords in the literature from 2020 to 2025 shows that "overfitting" has the highest frequency (350), followed by "regularization" (340), "dropout" (320), "data augmentation" (275), and "SAM" (270). Keywords related to optimization, such as "sharpness-aware," "optimizer," "Adam," and "SGD," also show relatively high frequencies. This pattern indicates that image classification research largely focuses on controlling overfitting through regularization and data augmentation, while also improving the optimization process and convergence. The emergence of approaches such as SAM, sharpness-aware optimization, MixUp, self-supervised learning, and Vision Transformer (ViT) indicates an increasing diversity of strategies for improving model generalization. These frequencies represent the number of times each keyword appears across the entire corpus; thus, a single article may contribute to more than one occurrence.

In terms of datasets, the frequency indicates the number of times a dataset is used or mentioned across all studies, not the number of articles that use it. ImageNet (250) and CIFAR-10 (200) are the most frequently used datasets, followed by CIFAR-100 (180), MNIST (150), Fashion-MNIST (130), and STL-10 (120). The dominance of these benchmark datasets indicates a trend in research toward using widely established datasets to compare performance, model robustness, and the effectiveness of overfitting mitigation strategies. Although more specialized datasets such as EuroSAT, Rice, Pascal VOC, and COCO are also used, reliance on a limited number of benchmark datasets has the potential to introduce benchmark bias. Therefore, the generalization ability of overfitting mitigation strategies on datasets with different characteristics, complexities, and distributions still requires further study.

2. Core Issues Underlying Overfitting in Image Classification

Before addressing RQ1, RQ2, and RQ3, the thematic analysis of the reviewed studies identified three recurring factors associated with overfitting in image classification tasks: limited labeled data, architectural complexity, and data and label quality. These factors were consistently reported across the literature and were found to influence the choice of mitigation techniques, optimization strategies, and training parameter settings. The following

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

discussion outlines the main findings related to these underlying contributors to overfitting and establishes the conceptual basis for the subsequent analysis.

Limited labeled data is the most common issue and directly leads to overfitting, especially in modern CNN- and Transformer-based models with high parameter capacity. Various studies, such as [34], [35], [36], [37], consistently show that models tend to memorise when the dataset is small, unbalanced, or when the annotation process is complicated. This condition is common in the medical domain [7], [38] and the industrial sector [39] which naturally have limited data space. The literature also confirms that without interventions such as data augmentation, transfer learning, or strong regularisation, models with large capacities will struggle to maintain reliable performance on unseen samples.

Large-parameter or very deep architectures have been shown to increase the risk of overfitting if not balanced with sufficient data. Models such as large-scale ResNets, EfficientNets, and Vision Transformers require strict regulatory mechanisms to prevent the memorisation of irrelevant features [6], [10], [13], [40]. Some studies have even found that models with more than 50 layers can experience a decline in validation accuracy on medium-sized datasets [6], [31]. Conversely, lightweight architectures such as MobileNet or small CNNs exhibit better training stability on limited datasets [9]. These findings confirm that architecture selection should not only follow model-complexity trends but also consider dataset characteristics.

Data quality plays a crucial role in determining training stability, model robustness, and prediction consistency. Visual noise, image artifacts, label inconsistencies, and annotation errors can cause the model to learn irrelevant patterns [6], [10], [12], [15], [41], [42]. Annotation errors can also produce unstable gradients and degrade validation performance [14], [15], while background noise and uncontrolled lighting can influence the model when inter-class variation is low [6]. Therefore, mitigating overfitting must be accompanied by improvements in data quality through dataset curation, label cleaning, and noise filtering.

In addition to data quality, knowledge distillation and feature distillation are widely used to address data limitations and architectural complexity by transferring informative representations from a high-capacity teacher model to a more compact student model. This approach can help models learn meaningful representations, reduce the tendency to memorize training data, and improve performance on unseen samples [16], [17], [18], [19]. Aligning feature representations across layers has also been reported to improve representation quality and reduce sensitivity to noisy inputs [17], [18], while simultaneously reducing parameter redundancy and improving learning efficiency under conditions of limited labeled data [26], [28].

3. Thematic Analysis Based on RQ

Table 2 presents a systematic mapping between the research questions (RQs), the methodological categories identified in the literature, and representative studies illustrating each approach. Due to space limitations, only selected references are shown to exemplify the methodological categories. At the same time, the complete corpus of 174 reviewed studies was considered in the broader analytical synthesis of this systematic literature review. This mapping provides an overview of how different approaches are used to mitigate overfitting, improve training convergence, and configure model parameters in deep learning–based image classification tasks. The mapping was developed using the predefined methodological coding framework described in the Materials and Methods section. During the coding process, a study could be assigned to multiple methodological themes and research questions whenever more than one relevant approach was reported.

Table 2. Systematic Mapping Between Research Questions (RQs) and Representative Studies from the Reviewed Literature

RQ	Methods	Key Reference
RQ1	Conventional Data Augmentation (geometric & photometric transformations)	[9], [22], [23]
	Advanced Augmentation Strategies (elastic transforms, CLAHE, MixUp/CutMix variants)	[7], [39], [43], [44], [45]
	Transfer Learning with Pretrained Models (feature reuse, fine-tuning)	[7], [24], [25], [39], [46]
	Semi-Supervised Learning (pseudo-labeling, consistency regularization)	[1], [9], [15], [17]
	GAN-Based Synthetic Data Generation	[21], [28], [33], [37], [47], [48], [49], [50], [51]

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

RQ	Methods	Key Reference
	Architectural Regularization: Dropout	[14], [28], [48]
	Architectural Regularization: Weight Decay (L2)	[10], [14]
	Architectural Regularization: Batch Normalization	[14], [52]
	Label Smoothing & Robust Loss Functions	[14], [15], [17], [18], [28]
	Image Quality Enhancement (denoising, filtering, histogram equalization)	[33], [40], [53], [54], [55], [56], [57]
	Knowledge & Feature Distillation	[16], [17], [18], [19], [20]
RQ2	Adaptive Optimization Algorithms (Adam, AdamW, SGD-Momentum)	[1], [6], [10], [31], [37], [48], [58], [59]
	Learning-Rate Calibration and Sensitivity Control	[6], [10], [31], [37].
	Dynamic Learning-Rate Scheduling (cosine annealing, warmup, decay policies)	[6], [10], [31], [37], [48]
	Batch Normalization as a Convergence Stabilizer	[6], [10], [14], [52], [60]
	Gradient Stabilization and Loss-Smoothing Techniques	[17], [28], [48], [61], [62]
RQ3	Learning-Rate Configuration (initial LR, LR scaling)	[1], [6], [10], [31], [37], [48]
	Batch-Size Adjustment (gradient variance control)	[21], [33], [40], [47], [49], [63]
	Epoch Scheduling and Training Duration Control	[1], [10], [31], [33], [37], [40], [63]
	Dropout-Rate Optimization	[1], [10], [31], [40], [48]
	Weight-Decay Regularization (L2)	[1], [6], [10], [31], [37], [48]
	Kernel / Filter-Size Configuration (spatial receptive-field control)	[33], [40], [47], [49], [63]

Source: (Budiman et al, 2025)

To strengthen the analytical synthesis, a frequency analysis of methodological categories was conducted on the 174 studies reviewed. Feature representation learning was the largest category (26.4%), followed by few-shot and semi-supervised learning (12.1%), and data augmentation (10.3%). Architectural regularization techniques, such as dropout and batch normalization, accounted for approximately 8.0% of the studies. Other approaches, including knowledge distillation, GAN-based data generation, and optimization strategies, appeared with lower frequency but indicated emerging trends in deep learning research. This distribution suggests that efforts to mitigate overfitting are increasingly focused on improving feature representation and data efficiency, rather than solely on conventional regularization.

For the frequency analysis, each study was classified once based on its most dominant methodological contribution, as assessed by research objectives, methods, experimental design, and main findings. This approach was used to avoid double counting, even though a single study might employ multiple techniques simultaneously. The classification results show that Feature Learning and Representation Learning is the largest category (46 studies or 26.4%), followed by Semi-Supervised and Few-Shot Learning (21 or 12.1%), Data Augmentation and Dataset Expansion (18 or 10.3%), Lightweight Architecture and Model Compression (16 or 9.2%), and Architectural Regularization such as Dropout and Batch Normalization (14 or 8.0%). Other categories include Other CNN Architecture Improvements (10 or 5.7%), Optimization and Learning Rate Strategies (9 or 5.2%), Robust Loss and Noise Handling and Domain Adaptation and Domain Generalization (8 or 4.6% each), GAN-based Data Generation (7 or 4.0%), Knowledge Distillation (6 or 3.4%), and Transfer Learning and Knowledge Transfer (11 or 6.3%).

This distribution indicates that model generalization research tends to involve a combination of several complementary approaches, so the frequency of categories represents the primary methodological focus rather than the total number of techniques used. Some methods, such as dropout, batch normalization, weight decay, and learning rate scheduling, may appear in more than one research question because they serve different functions: RQ1 examines their role in mitigating overfitting, RQ2 in optimization convergence and stability, and RQ3 in parameter configurations that affect generalization. Thus, the appearance of the same method in multiple RQs reflects the multidimensional nature of the deep learning training process, rather than a repetition of the analysis

RQ1: How are methods used in the literature to overcome overfitting in image classification tasks?

Thematic analysis shows that overfitting in image classification is primarily influenced by limitations in labeled data, the complexity of the model's architecture and capacity, and the quality of the data or labels. These factors form the basis for the development of an integrated taxonomy in this study. The mitigation strategies identified include data augmentation, transfer learning, regularization, optimization, and knowledge or feature distillation, which collectively illustrate the relationship between the causes of overfitting and their mitigation mechanisms in the reviewed literature.

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

Data augmentation is the most widely used strategy because it can expand the data distribution without requiring additional labeling. For small datasets or those with limited visual diversity, augmentation helps expand the representation space and improve the model's robustness to data variations. Common techniques such as rotation, flipping, random cropping, and color jitter have been used to enhance robustness against spatial and lighting variations [9], [22], [23]. Some studies have also applied more complex transformations, such as elastic deformation and contrast-limited adaptive histogram equalization (CLAHE), to improve image quality before the training process [7], [39]. Conceptually, augmentation encourages the model to learn features that are more invariant to small changes in the input distribution. This approach is particularly important in the medical and agricultural domains, which generally face limitations in the availability of large-scale datasets.

Transfer learning is a crucial strategy when the target dataset is too small to train from scratch. Pre-trained models, such as those on ImageNet, can provide stable, information-rich initial representations [7], [24], [25], [39], [46]. This method accelerates convergence while reducing the risk of overfitting, as the initial weights already include tested generic features. Fine-tuning then allows adaptation to new domains with minimal changes to the model structure. Empirically, this approach is efficient when classes have complex visual structures or high intra-class variation.

When labeled data is limited but unlabeled data is abundant, the semi-supervised learning approach is widely used. Pseudo-labelling uses model predictions as temporary labels to expand the labeled dataset and to regularise consistency [1], [9], [15], [17]. Consistency regularisation ensures that the model produces stable predictions across a range of input transformations. This approach has been shown to help reduce overfitting by enriching the training data distribution.

Generative Adversarial Networks (GAN)-based techniques are used to generate synthetic data that resembles the original distribution. Studies show that GANs are very useful for addressing class imbalance and increasing image variation in specialised domains such as medicine [21], [28], [33], [37], [47], [48], [49]. The main advantage of GANs is their ability to generate complex visual patterns that are difficult to produce with conventional augmentation methods. However, several studies highlight that the quality of synthetic images is highly dependent on the stability of GAN training.

Regularisation is a fundamental approach to controlling model complexity and preventing overfitting. One of the most widely used techniques is dropout, which randomly deactivates neurons during training. This approach reduces co-adaptation between neurons and makes the model more robust to data variation [14], [28], [48]. In addition, weight decay (L2 regularisation) plays an important role in pushing network weights towards smaller values, thereby reducing the risk of over-reliance on certain features [10], [14]. Another widely used technique is Batch Normalisation (BN), which stabilises the activation distribution and accelerates convergence. BN is known as an implicit regulariser because it smooths the loss landscape, thereby reducing gradient volatility and overfitting [14], [52]. The architectural regularisation approach has proven to be a significant contribution, especially for deep models with large numbers of parameters, such as ResNet, DenseNet, and EfficientNet. These techniques help maintain training stability while improve performance beyond the training data.

When a dataset contains inaccurate or inconsistent labels, label smoothing is one of the most effective techniques for mitigating the impact of annotation errors. This technique effectively reduces the model's overconfidence, preventing it from easily learning noise from unclean labels [14], [15], [17], [18], [28]. By spreading the target distribution more evenly, label smoothing helps the model build a more stable representation that is resistant to annotation errors. In addition, several studies have adopted robust loss functions, such as generalised cross-entropy or symmetric cross-entropy, which are designed to reduce excessive penalties when models make incorrect predictions due to label inaccuracies. This approach provides greater tolerance to label noise and minimises the risk of models learning incorrect patterns that do not reflect the actual data distribution.

Various techniques have been proposed to improve image quality, ensuring that visual information structures are in optimal condition before the training process begins [33], [40], [63]. Methods such as filtering, histogram equalisation, and denoising are used to remove artefacts and noise that could interfere with feature learning. This approach has proven very effective, especially on real-world datasets collected under inconsistent lighting or environmental conditions. By improving image quality during pre-processing, the model obtains cleaner signal information, thereby reducing the risk of overfitting and improving robustness across varying image conditions.

Distillation techniques are used to obtain lighter models while maintaining predictive performance. Knowledge distillation transfers representations or prediction distributions from a high-capacity model (the "teacher") to a "student" model, allowing the smaller model to learn important features without replicating the full complexity of the source model [16], [17], [18], [19], [20]. Meanwhile, feature distillation condenses internal representations at an intermediate level to reduce redundancy and form a more structured feature space. In general, distilled models are reported to have more stable performance and a lower tendency toward overfitting compared to models trained from scratch.

The comparison results show that there is no single overfitting mitigation strategy that is universally superior. The choice depends on the cause of overfitting, the characteristics of the dataset, the model's complexity, and the application's objectives. In general, these strategies can be grouped into four approaches: first, data augmentation and synthetic data to address data limitations; second, transfer learning and semi-supervised learning to leverage trained representations and unlabeled data; third, architectural regularization and robust loss functions to control model capacity and resilience to noise; and fourth, knowledge distillation to produce lighter models that retain good generalization capabilities.

RQ2: How can methods improve convergence during the image classification model training process?

In RQ2, the analysis focuses on the impact of various methods on optimization dynamics, particularly in accelerating convergence, improving training stability, and enhancing the efficiency of the optimization process. The results of the study show that convergence is a critical concern in image classification, particularly for CNNs, EfficientNet, and Vision Transformers, which face a non-smooth loss landscape, gradient fluctuations, and an unbalanced data distribution. These conditions can increase the sensitivity of the optimization process and the risk of divergence. The methods discussed in RQ1 were then re-examined based on their contribution to convergence.

Based on a systematic mapping of the analyzed articles, methods for improving convergence can be grouped into four main approaches: (1) selection of optimizers and adaptive optimization strategies, (2) adjustment and calibration of the learning rate, (3) use of dynamic learning rate schedulers, and (4) internal stabilization mechanisms such as batch normalization. These four approaches form the basis for improving the stability of the training process and the model's robustness to unseen data. A further discussion of each approach is presented in the following sections based on relevant literature.

The choice of optimizer plays a crucial role in determining the speed and stability of model convergence. The Adam, AdamW, and SGD with momentum algorithms generally exhibit faster and more stable convergence compared to standard SGD [1], [6], [10], [31], [37], [48]. Adam and AdamW utilize first- and second-order moment estimates to adjust the learning rate based on gradient information, whereas SGD with momentum helps dampen gradient oscillations and results in a more stable optimization process, particularly on data with high levels of noise or intra-class variation. However, these advantages do not always translate to equivalent out-of-sample performance. Adaptive optimizers tend to speed up the training process, while SGD+Momentum has been shown in several studies to provide better generalization by guiding the model toward a flatter loss landscape. Thus, the choice of optimizer must consider the balance between convergence speed and prediction quality on unseen data.

The learning rate is one of the parameters that most significantly influences the dynamics of weight updates during training. A value that is too high can cause overshooting and optimization instability, while a value that is too low slows convergence and increases the risk of the model getting stuck on a plateau. Therefore, proper adjustment of the learning rate plays a crucial role in maintaining the speed and stability of convergence [6], [10], [31], [37]. In transfer learning, using a smaller learning rate during fine-tuning can result in more stable convergence because the initial representation of the pre-trained model does not undergo excessive changes [6], [10]. Thus, the selection of the learning rate is not only a decision made in the early stages of training but also a critical component in maintaining the effectiveness and stability of the optimization process.

In addition to setting an initial learning rate, several studies have used learning rate schedulers such as cosine annealing, step decay, exponential decay, and warmup to stabilize the optimization process [6], [10], [31], [37], [48]. Warmup helps reduce gradient instability in the early stages of training, while cosine annealing gradually decreases the learning rate, thereby supporting a more stable convergence process. Several studies

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

have also shown that learning rate schedulers can act as implicit regularization by helping the model avoid sharp minima that hinder generalization. Thus, dynamic learning rate adjustment not only improves convergence but also enhances training stability and the predictive robustness of image classification models.

Batch Normalization (BN) helps accelerate training while stabilizing model convergence by controlling activation variance and gradient stability [6], [10], [14], [52]. BN normalizes the input in each batch based on its mean and variance, thereby making the optimization process more stable. In large architectures such as ResNet and DenseNet, BN helps reduce the risk of exploding and vanishing gradients, especially in very deep networks. In addition to improving training efficiency, BN can also provide an implicit regularization effect by helping to smooth the loss landscape and guide the optimizer toward a more stable state. Thus, the application of BN contributes to the stability of the training process and the model's robustness when faced with different data distributions.

Several studies have used gradient clipping, consistency regularization, loss smoothing, and interlayer normalization to stabilize gradient dynamics and prevent excessive fluctuations during optimization [14], [17], [28], [48]. Gradient clipping helps control weight updates when gradients are extreme, while consistency regularization and loss smoothing promote a smoother loss surface, leading to more stable optimization. Findings indicate a trade-off in convergence strategies. Adam and AdamW tend to accelerate early convergence, whereas momentum-based SGD can provide better generalization after convergence. Aggressive learning rate scheduling can accelerate optimization but increases sensitivity to hyperparameters. Batch normalization generally improves stability, although its effectiveness may decrease with very small batch sizes. Thus, the evaluation of convergence strategies must consider not only the speed of optimization but also training stability and model generalization ability.

A systematic review shows that model convergence does not depend on a single technique, but rather on a combination of complementary strategies. The use of adaptive optimizers provides a stable initial foundation, while learning rate adjustments and the application of schedulers help regulate the dynamics of weight updates throughout training. Batch Normalisation and various gradient stabilization techniques help maintain balanced learning across all layers of the model. This set of methods forms a comprehensive framework for improving convergence, ensuring the training process is more effective, efficient, and stable, and supporting more reliable performance on unseen data.

RQ3: How can parameters reduce overfitting while maintaining or improving image classification model performance?

RQ3 examines the impact of training parameter configurations on the performance and generalization ability of image classification models. The review identifies six key parameters—learning rate, batch size, number of epochs, dropout rate, weight decay, and kernel/filter size—which collectively play a role in controlling gradient dynamics, model capacity, and feature representation stability to reduce overfitting.

The learning rate is one of the most critical and frequently discussed parameters in the literature because it serves as the central controller in weight updates during training. The selection of an appropriate learning rate determines whether the model converges stably or diverges [1], [6], [10], [37], [48]. Several studies have also shown that pre-trained models are more sensitive to the learning rate, so a relatively conservative value is needed to prevent excessive changes to the initial representation. Thus, the learning rate serves as the primary control mechanism that influences learning dynamics, training stability, and the model's ability to maintain performance on data outside the training distribution.

The batch size determines the number of samples processed in each iteration and affects the characteristics of the gradient. A small batch size can act as a form of natural regularization because it produces gradient variations that help the model avoid sharp minima associated with poor generalization performance [21], [40], [49]. However, a batch size that is too small can increase training instability, while a large batch size produces smoother gradients and accelerates convergence but may increase the risk of overfitting. Thus, the choice of batch size should be tailored to the complexity of the architecture and the size of the dataset, with small to medium sizes generally providing a balance between training stability and the model's predictive robustness.

The number of epochs determines the duration of the training process and plays a role in controlling the risk of overfitting. Training that lasts too long can cause the model to learn noise or specific patterns in the training data [1], [10], [31], [33], [37], [40], [63], whereas too few epochs can hinder the formation of an optimal feature representation. Therefore, the number of epochs is generally combined with early stopping or learning rate decay so that training is stopped under appropriate conditions and the memorization phase can be avoided. The literature indicates that the number of epochs needs to be determined adaptively based on validation performance to maintain a balance between feature exploration and overfitting mitigation.

Dropout is one of the most commonly used regularisation mechanisms to reduce co-adaptation between neurons. This technique consistently improves representation robustness by forcing the network to build more robust representations [1], [10], [31], [40], [48], [63]. Dropout works by randomly deactivating units during training, so that the model does not rely on specific activation paths and is encouraged to learn more general patterns. However, dropout effectiveness depends heavily on the chosen dropout rate. These studies report that excessively high dropout values can eliminate important information and negatively affect accuracy, whereas excessively low values are less effective at mitigating overfitting. The most recommended range is 0.2–0.5, which is considered to provide the best balance between regularisation and information retention, depending on the architecture's depth and the dataset's complexity.

Weight decay, or L2 regularization, is an important parameter for limiting weight growth and controlling model complexity. Its application helps stabilize gradient updates and reduces the model's tendency to learn features that are too specific to the training data [1], [6], [10], [37], [48]. such as AdamW because weight regularization is separated from the gradient adaptation mechanism, making the optimization process more stable and consistent. This strategy is particularly relevant for deep models and datasets of limited size that are at higher risk of overfitting.

The kernel or filter size affects spatial coverage, model capacity, and training stability. Larger kernels can capture a broader spatial context but increase the number of parameters and the risk of overfitting, especially on limited datasets [33], [40], [47], [49], [63]. Conversely, small kernels, particularly 3×3 kernels, tend to be more efficient and stable. Thus, adjusting the kernel size can be used to control the architecture's capacity. The effectiveness of these parameters also depends on the model configuration and dataset characteristics; therefore, optimization must be conducted flexibly and contextually, rather than based on a single configuration considered universally optimal.

A systematic review shows that training parameters play a strategic role in controlling overfitting and form an important foundation in image classification experimental design. Six main parameters—learning rate, batch size, number of epochs, dropout rate, weight decay, and kernel and filter size—form a complementary regulatory framework in controlling gradient sensitivity, representation capacity, and learning process stability. The accuracy of each parameter adjustment directly contributes to the model's ability to maintain reliable performance beyond the training data. Thus, training parameter settings are not merely a technical aspect but a key component in the design of modern image classification systems aiming for stable, efficient, and well-generalised performance.

B. DISCUSSION

A cross-study synthesis shows that overfitting in deep learning-based image classification is influenced by limitations in labeled data, architectural complexity, data and label quality, and the effectiveness of representation transfer mechanisms such as knowledge distillation. The strategies examined include regularization, optimization, and architectural mechanisms to improve generalization and training stability. Findings related to overparameterization and double descent also indicate that model size alone is insufficient to explain the occurrence of overfitting, as high-capacity models can still produce good predictive performance when supported by optimal conditions and sufficient data availability.

In RQ1, data-driven strategies such as data augmentation, transfer learning, and synthetic data generation are widely used to address data limitations and variability. Conventional augmentations, such as rotation, flipping, and color jitter, have been shown to improve model robustness, particularly on small datasets [9], [22], [23]. However, augmentation alone is not always sufficient when the model has high capacity or when label noise is present; thus, transfer learning from pretrained models can provide more stable feature representations [7], [24], [25], [39], [46]. In RQ2, the choice of optimizer affects convergence dynamics [1], [6], [10], [31], [37], [48], while learning rate calibration and dynamic scheduling help control gradient behavior [6], [10], [31], [37]. Although Adam and AdamW tend to accelerate convergence, faster convergence does not always result in better predictive robustness. Meanwhile, momentum-based SGD can produce more stable solutions by converging toward flatter minima. In RQ3, parameter configurations such as the learning rate [1], [6], [10], [31], [37], [48], batch size [21], [33], [40], [47], [49], [63], weight decay, and kernel size

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

[33], [40], [47], [49], [63] play a role in controlling gradient sensitivity, gradient variance, model capacity, and the tendency to learn overly specific patterns. Overall, effective overfitting mitigation is largely determined by a combination of three complementary dimensions: data-level strategies, architectural regularization, and optimization control. Together, these three elements help broaden the learning distribution, control model capacity, and stabilize training dynamics so that generalization can be maintained across various image classification scenarios.

Technically, the SLR results show that no single method is consistently capable of addressing overfitting. Mitigation effectiveness is achieved more through a combination of complementary strategies. Intensive data augmentation [1], [40], [63], [64] can be reinforced with regularization techniques such as dropout and batch normalization [10], [31], while transfer learning tends to be more stable when fine-tuning is performed with a low learning rate [6]. These findings suggest that a model's robustness to unseen data is influenced by the balance between parameter complexity, data variability, and gradient stability during training.

From a scientific perspective, this SLR indicates a shift in the focus of image classification research from increasing architectural capacity toward representation stability and parameter efficiency. Studies on knowledge distillation [32], [37], [48] show that lighter-weight models can achieve performance equivalent to or even better than large models through more effective redundancy reduction and representation compression. Thus, prediction reliability is not solely determined by model size, but also by the ability to design informative, efficient, and well-controlled representations.

The main strength of the analyzed studies lies in the diversity of their application contexts, covering the medical field [6], [8], [10], [12], [26], agriculture [65], and industry and manufacturing [26], [31], [52]. Several studies also conducted cross-architecture and cross-dataset evaluations [10], [31], [66], thereby providing insight into the consistency of the methods' performance under different conditions. The use of approaches such as GANs [32], knowledge distillation [19], [48], and dynamic schedulers [6], [10], [21], [28], [37], [47], [48] demonstrates an increasingly diverse methodological development.

On the other hand, these studies still have limitations, particularly their reliance on popular datasets such as CIFAR-10, CIFAR-100, and ImageNet, which can limit generalization to domains with more complex visual characteristics [31], [52]. Research on the impact of heavy noise, annotation errors, long-term stability, and continual learning is also still limited [6], [14], [48], [65]. From a review perspective, the use of Scopus as the sole source and relatively narrow keywords may result in relevant studies from other databases or using different terminology going unidentified. Although a PRISMA-based selection process was applied, future research could expand the database sources and search strategies to improve literature coverage.

These SLR findings are relevant to the application of image classification models in various real-world environments. In the medical field, domain-specific augmentation, transfer learning, and distillation have the potential to improve the accuracy of diagnostic systems without requiring large datasets [12], [52], [67]. Meanwhile, in the agricultural and industrial sectors, distillation and low-computational-cost optimization support the deployment of models on edge devices with limited resources [14], [17], while dynamic schedulers and momentum-based optimizers can improve the efficiency of computational resource utilization while maintaining training stability [10], [1], [6]. Overall, the results of this systematic literature review indicate that overfitting mitigation strategies must simultaneously account for data limitations, hardware capabilities, and application domain characteristics to ensure effective deployment in real-world environments.

Corpus analysis shows that although there has been progress in mitigating overfitting and stabilizing convergence, there are still research gaps in four main dimensions: models, datasets, methodologies, and implementation. Previous research has been dominated by large architectures such as ResNet, DenseNet, EfficientNet, and Vision Transformer [6], [10], [31], while the development of lightweight models for devices with limited resources remains relatively limited [68], [33]. Evaluations of memory, latency, and power consumption have also been scarce, as has multimodal integration—such as combining RGB images with multispectral data, sensor metadata, or physiological signals [65], [69]. In terms of datasets, reliance on CIFAR-10, CIFAR-100, MNIST, and ImageNet does not fully represent real-world conditions [1], [14], [19], [20], [31]. Data limitations, class imbalance, privacy concerns, as well as variations in lighting, noise, occlusion, and geographical conditions further increase the risk of overfitting, particularly in the medical, agricultural, and manufacturing domains [6], [8], [12]. Furthermore, variations in lighting, noise, occlusion, and geographic conditions are still limited in many datasets, which can reduce a model's generalization ability [6], [10]. These conditions highlight the need for datasets that are more representative of real-world operational variations to improve model robustness.

In terms of methodology, there are currently no consistent benchmark standards or evaluation protocols, while differences in training configurations and hyperparameters make it difficult to compare results or conduct meta-analyses [10], [31]. Explainability remains limited, particularly in medical [6], [52] and industrial [65] domains. Testing in real-world edge environments and under dynamic conditions—such as changes in lighting, moving objects, and sensor signal degradation—remains incomplete [40]. Therefore, future research should focus on lightweight models, more representative datasets, consistent benchmarks, multimodal integration, interpretability, edge computing efficiency, and robustness under real-time conditions. Furthermore, the use of a single database and the absence of study quality assessments, risk of bias, and inter-reviewer agreement constitute limitations of this systematic literature review (SLR), future research is recommended to utilize multiple databases and involve more than one reviewer, as well as quality assessment tools, to enhance the completeness and reliability of the synthesis.

CONCLUSION

This study presents a Systematic Literature Review (SLR) on strategies for addressing overfitting, improving convergence, and optimizing training parameters in deep learning-based image classification. The results of the review indicate that overfitting remains a challenge, particularly under conditions of limited training data, complex network architectures, and varying data quality. Performance improvements generally do not depend on a single technique but rather on a combination of complementary methods, such as data augmentation, transfer learning, regularization, adaptive optimization, and the selection of appropriate training parameters. Knowledge distillation is also increasingly being used to reduce computational requirements while maintaining predictive performance.

The main contribution of this research is the development of an analytical taxonomy that links the causes of overfitting to mitigation strategies and training mechanisms found in the literature. The study also identifies the need for further research regarding the development of more representative datasets, improved annotation quality, computationally efficient architectures, and more consistent evaluation procedures. Overall, the resulting taxonomy can serve as a reference for researchers and practitioners in selecting approaches to build image classification models that are more robust, efficient, and adaptable to real-world application needs.

REFERENCE

[1] F. Feng, M. Gao, R. Liu, S. Yao, and G. Yang, 'A deep learning framework for crop mapping with reconstructed Sentinel-2 time series images', *Comput. Electron. Agric.*, vol. 213, 2023, doi: 10.1016/j.compag.2023.108227.

[2] Y. Asam and Z. Zhou, 'A dual-path feature learning neural network for enhancing image classification in marine biology', *Multimedia Tools Appl*, 2025, doi: 10.1007/s11042-025-21010-x.

[3] J. Qiu *et al.*, 'A Flattened-Tree Genetic Programming Approach with Multi-Scale Feature Extraction for Image Classification', *Memetic Comput.*, vol. 17, no. 3, 2025, doi: 10.1007/s12293-025-00472-4.

[4] H. Song, 'A Leading but Simple Classification Method for Remote Sensing Images', *Ann. Emer. Tech. Comp.*, vol. 7, no. 3, pp. 1–20, 2023, doi: 10.33166/AETiC.2023.03.001.

[5] E. Keskin Bilgiç, İ. Gökbay, and Y. Kayar, 'Innovative Approaches to Clinical Diagnosis: Transfer Learning in Facial Image Classification for Celiac Disease Identification', *Appl. Sci.*, vol. 14, no. 14, 2024, doi: 10.3390/app14146207.

[6] Y. Suo, Z. He, and Y. Liu, 'Deep learning CS-ResNet-101 model for diabetic retinopathy classification', *Biomed. Signal Process. Control*, vol. 97, 2024, doi: 10.1016/j.bspc.2024.106661.

[7] Z. He *et al.*, 'Deconv-transformer (DecT): A histopathological image classification model for breast cancer based on color deconvolution and transformer architecture', *Inf Sci*, vol. 608, pp. 1093–1112, 2022, doi: 10.1016/j.ins.2022.06.091.

[8] S.-Y. Lu, Z. Zhu, Y. Tang, X. Zhang, and X. Liu, 'CTBViT: A novel ViT for tuberculosis classification with efficient block and randomized classifier', *Biomed. Signal Process. Control*, vol. 100, 2025, doi: 10.1016/j.bspc.2024.106981.

[9] X. Tong, Z. Liang, and F. Liu, 'Succulent Plant Image Classification Based on Lightweight

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

GoogLeNet with CBAM Attention Mechanism', *Appl. Sci.*, vol. 15, no. 7, 2025, doi: 10.3390/app15073730.
[10] Y. Hu, J. Tang, Y. Xu, R. Xu, and B. Huang, 'EL-DenseNet: a novel method for identifying the flame state of converter steelmaking based on dense convolutional neural networks', *Signal Image Video Process.*, vol. 18, no. 4, pp. 3445–3457, 2024, doi: 10.1007/s11760-024-03011-9.
[11] M. Piao, Y. Sheng, J. Yan, and C. H. Jin, 'Image Hash Layer Triggered CNN Framework for Wafer Map Failure Pattern Retrieval and Classification', *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 4, 2024, doi: 10.1145/3638053.
[12] Y. Kong, X. Ma, and C. Wen, 'A New Method of Deep Convolutional Neural Network Image Classification Based on Knowledge Transfer in Small Label Sample Environment', *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22030898.
[13] F. Xu, P. Wang, and H. Xu, 'Deep pyramidal residual networks with inception sub-structure in image classification', *J. Intelligent Fuzzy Syst.*, vol. 45, no. 4, pp. 5885–5906, 2023, doi: 10.3233/JIFS-230569.
[14] Q. Huang, H. Zhang, M. Xue, J. Song, and M. Song, 'A Survey of Deep Learning for Low-shot Object Detection', *ACM Comput Surv*, vol. 56, no. 5, 2024, doi: 10.1145/3626312.
[15] D. Zhai, R. Shi, J. Jiang, and X. Liu, 'Rectified Meta-learning from Noisy Labels for Robust Image-based Plant Disease Classification', *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 18, no. 1s, 2022, doi: 10.1145/3472809.
[16] S. Li, H. Hu, S. Huo, and H. Liang, 'Clean, performance-robust, and performance-sensitive historical information based adversarial self-distillation', *IET Comput. Vision*, vol. 18, no. 5, pp. 591–612, 2024, doi: 10.1049/cvi2.12265.
[17] M. Kang *et al.*, 'Efficient one-shot federated learning on medical data using knowledge distillation with image synthesis and client model adaptation', *Med. Image Anal.*, vol. 105, 2025, doi: 10.1016/j.media.2025.103714.
[18] M. Liu, Y. Yu, Z. Ji, J. Han, and Z. Zhang, 'Tolerant Self-Distillation for image classification', *Neural Netw.*, vol. 174, 2024, doi: 10.1016/j.neunet.2024.106215.
[19] Y. Zhou, Z. Wang, and J. Li, 'Knowledge Distillation Based on Narrow-Deep Networks', *Neural Process Letters*, vol. 56, no. 3, 2024, doi: 10.1007/s11063-024-11646-5.
[20] J. Zhang, Z. Chen, and L. Dai, 'Unleashing the Power of Each Distilled Image', *IEEE Trans Image Process*, 2025, doi: 10.1109/TIP.2025.3624626.
[21] J. Mi, C. Ma, L. Zheng, M. Zhang, M. Li, and M. Wang, 'WGAN-CL: A Wasserstein GAN with confidence loss for small-sample augmentation', *Expert Sys Appl*, vol. 233, 2023, doi: 10.1016/j.eswa.2023.120943.
[22] K. Alomar, H. I. Aysel, and X. Cai, 'Data Augmentation in Classification and Segmentation: A Survey and New Strategies', *J. Imaging*, vol. 9, no. 2, 2023, doi: 10.3390/jimaging9020046.
[23] Z. Wang, Y. Guo, Q. Li, G. Yang, and W. Zuo, 'DualAug: Exploiting additional heavy augmentation with OOD data rejection', *Neurocomputing*, vol. 654, 2025, doi: 10.1016/j.neucom.2025.131209.
[24] Y. Chen, X. Fang, and H. Han, 'An improved transfer learning algorithm using integration method and its application in image segmentation and recognition', *Multimedia Tools Appl*, vol. 84, no. 24, pp. 28085–28114, 2025, doi: 10.1007/s11042-024-20290-z.
[25] J. Su, X. Yu, X. Wang, Z. Wang, and G. Chao, 'Enhanced transfer learning with data augmentation', *Eng Appl Artif Intell*, vol. 129, 2024, doi: 10.1016/j.engappai.2023.107602.
[26] M. Sharma, J. Heard, E. Saber, and P. Markopoulos, 'Convolutional Neural Network Compression via Dynamic Parameter Rank Pruning', *IEEE Access*, vol. 13, pp. 18441–18456, 2025, doi: 10.1109/ACCESS.2025.3533419.
[27] W. Wei, X. Fu, S. Ma, Y. Zhu, and N. Lu, 'Reducing overfitting in vehicle recognition by decorrelated sparse representation regularisation', *IET Comput. Vision*, vol. 18, no. 8, pp. 1351–1361, 2024, doi: 10.1049/cvi2.12320.
[28] Z. Ren, Y.-D. Zhang, and S. Wang, 'A Hybrid Framework for Lung Cancer Classification', *Electronics (Switzerland)*, vol. 11, no. 10, 2022, doi: 10.3390/electronics11101614.
[29] H. Sun *et al.*, 'An Improved Adam's Algorithm for Stomach Image Classification', *Algorithms*, vol. 17, no. 7, 2024, doi: 10.3390/a17070272.
[30] Y. Zheng *et al.*, 'Cyclic Learning Rate-Based Co-Training for Image Classification With Noisy Labels', *IEEE Access*, vol. 13, pp. 6292–6305, 2025, doi: 10.1109/ACCESS.2025.3526332.
[31] L. Hao, K. Hao, B. Wei, and X.-S. Tang, 'Boosting the transferability of adversarial examples via stochastic serial attack', *Neural Netw.*, vol.

150, pp. 58–67, 2022, doi: 10.1016/j.neunet.2022.02.025.
[32] J. Hu *et al.*, 'APDL: an adaptive step size method for white-box adversarial attacks', *Complex Intell. Syst.*, vol. 11, no. 1, 2025, doi: 10.1007/s40747-024-01748-x.
[33] S. Q. Gilani and O. Marques, 'Skin lesion analysis using generative adversarial networks: a review', *Multimedia Tools Appl*, vol. 82, no. 19, pp. 30065–30106, 2023, doi: 10.1007/s11042-022-14267-z.
[34] B. Shi, W. Li, J. Huo, P. Zhu, L. Wang, and Y. Gao, 'Global- and local-aware feature augmentation with semantic orthogonality for few-shot image classification', *Pattern Recogn.*, vol. 142, 2023, doi: 10.1016/j.patcog.2023.109702.
[35] M. Momeny *et al.*, 'Greedy Autoaugment for classification of mycobacterium tuberculosis image via generalized deep CNN using mixed pooling based on minimum square rough entropy', *Comput. Biol. Med.*, vol. 141, 2022, doi: 10.1016/j.compbiomed.2021.105175.
[36] X. Zhang, C. Yuan, W. Sun, and S. K. Jha, 'Image Emotion Classification Network Based on Multilayer Attentional Interaction, Adaptive Feature Aggregation', *Comput. Mater. Continua*, vol. 75, no. 2, pp. 4273–4291, 2023, doi: 10.32604/cmc.2023.036975.
[37] C. P. Lau, J. Liu, H. Souri, W.-A. Lin, S. Feizi, and R. Chellappa, 'Interpolated Joint Space Adversarial Training for Robust and Generalizable Defenses', *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 11, pp. 13054–13067, 2023, doi: 10.1109/TPAMI.2023.3286772.
[38] Z.-Z. Wu *et al.*, 'Domain Adaptation via Feature Disentanglement for cross-domain image classification', *Appl. Soft Comput.*, vol. 172, 2025, doi: 10.1016/j.asoc.2025.112868.
[39] C. Li *et al.*, 'Domain generalization on medical imaging classification using episodic training with task augmentation', *Comput. Biol. Med.*, vol. 141, 2022, doi: 10.1016/j.compbiomed.2021.105144.
[40] Q. Han *et al.*, 'DM-CNN: Dynamic Multi-scale Convolutional Neural Network with uncertainty quantification for medical image classification', *Comput. Biol. Med.*, vol. 168, 2024, doi: 10.1016/j.compbiomed.2023.107758.
[41] S. Wang, H. Ben, Y. Hao, X. He, and M. Wang, 'Boosting Hyperspectral Image Classification with Dual Hierarchical Learning', *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 19, no. 1, 2023, doi: 10.1145/3522713.
[42] C. Ningthoujam, T. C. Chinghtham, B. Brahma, and A. K. Bhoi, 'Hybrid CNN-KNN Model for Image Annotation: Combining Deep Learning and Instance-Based Learning', *J. Inst. Eng. Ser. B*, 2025, doi: 10.1007/s40031-025-01239-8.
[43] L. Yan, Y. Ye, C. Wang, and Y. Sun, 'LocMix: local saliency-based data augmentation for image classification', *Signal Image Video Process.*, vol. 18, no. 2, pp. 1383–1392, 2024, doi: 10.1007/s11760-023-02852-0.
[44] X. Xiong *et al.*, 'VariMix: A variety-guided data mixing framework for explainable medical image classifications', *Computer Methods and Programs in Biomedicine*, vol. 271, p. 109016, Nov. 2025, doi: 10.1016/j.cmpb.2025.109016.
[45] T. Li *et al.*, 'Salient Features Guided Augmentation for Enhanced Deep Learning Classification in Hematoxylin and Eosin Images', *Comput. Mater. Continua*, vol. 84, no. 1, pp. 1711–1730, 2025, doi: 10.32604/cmc.2025.062489.
[46] K. Sun, Y. Yin, F. Dong, and X. Sun, 'Hyperspectral classification method based on M-ResHSDC', *Multimedia Tools Appl*, vol. 83, no. 16, pp. 49767–49785, 2024, doi: 10.1007/s11042-023-17515-y.
[47] B. Wang, X. Hu, C. Zhang, P. Li, and P. S. Yu, 'Hierarchical GAN-Tree and Bi-Directional Capsules for multi-label image classification', *Knowl Based Syst*, vol. 238, 2022, doi: 10.1016/j.knosys.2021.107882.
[48] B. Gholami, Q. Liu, M. El-Khamy, and J. Lee, 'Multiexpert Adversarial Regularization for Robust and Data-Efficient Deep Supervised Learning', *IEEE Access*, vol. 10, pp. 85080–85094, 2022, doi: 10.1109/ACCESS.2022.3196780.
[49] D. Song, Y. Tang, B. Wang, J. Zhang, and C. Yang, 'Two-Branch Generative Adversarial Network With Multiscale Connections for Hyperspectral Image Classification', *IEEE Access*, vol. 11, pp. 7336–7347, 2023, doi: 10.1109/ACCESS.2022.3232152.
[50] D. Mukherkjee, P. Saha, D. Kaplun, A. Sinitca, and R. Sarkar, 'Brain tumor image generation using an aggregation of GAN models with style transfer', *Sci Rep*, vol. 12, no. 1, p. 9141, Jun. 2022, doi: 10.1038/s41598-022-12646-y.
[51] Z. Liao, S. Hu, Y. Zhang, and Y. Xia, 'Unleashing the potential of open-set noisy samples against label noise for medical image classification', *Med. Image Anal.*, vol. 105, 2025, doi: 10.1016/j.media.2025.103702.
[52] C. Zhang, X. Zhang, and D. Tu, 'A Set of Comprehensive Evaluation System for

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7948

Different Data Augmentation Methods', *Mob. Inf. Sys.*, vol. 2022, 2022, doi: 10.1155/2022/8572852.

[53] S. Cheng, P. Li, K. Han, Y. Zheng, H. Xu, and Y. Yao, 'DMFP: Dynamic multiscale feature perturbations for transferable adversarial attacks', *Knowl Based Syst*, vol. 330, 2025, doi: 10.1016/j.knosys.2025.114469.

[54] P. Huang, Z. Yang, W. Wang, and F. Zhang, 'Denoising Low-Rank Discrimination based Least Squares Regression for image classification', *Inf Sci*, vol. 587, pp. 247–264, 2022, doi: 10.1016/j.ins.2021.12.031.

[55] Z. Yang, D. Wang, P. Huang, M. Wan, and G. Yang, 'Regularisation constrained denoising discriminant least squares regression for image classification', *Expert Sys Appl*, vol. 252, 2024, doi: 10.1016/j.eswa.2024.124253.

[56] J. Guo, H. Wang, X. Xue, M. Li, and Z. Ma, 'Real-time classification on oral ulcer images with residual network and image enhancement', *IET Image Proc.*, vol. 16, no. 3, pp. 641–646, 2022, doi: 10.1049/ipr2.12144.

[57] M. Aamir, Z. Rahman, W. Ahmed Abro, U. Aslam Bhatti, Z. Ahmed Dayo, and M. Muhammad, 'Brain tumor classification utilizing deep features derived from high-quality regions in MRI images', *Biomed. Signal Process. Control*, vol. 85, 2023, doi: 10.1016/j.bspc.2023.104988.

[58] S. Iqbal, A. N. Qureshi, K. Aurangzeb, M. Alhussein, S. I. Haider, and I. Rida, 'AMIAC: adaptive medical image analyzes and classification, a robust self-learning framework', *Neural Comput. Appl.*, vol. 37, no. 25, pp. 20451–20479, 2025, doi: 10.1007/s00521-023-09209-1.

[59] N. Abdel Samee, E. H. Houssein, E. Saber, G. Hu, and M. Wang, 'Integrated deep learning-based IRACE and convolutional neural networks for chest X-ray image classification', *Knowl Based Syst*, vol. 329, 2025, doi: 10.1016/j.knosys.2025.114293.

[60] D. Xue, J. Huang, R. Zhou, Y. Tai, and J. Zhang, 'Secured COVID-19 CT image classification based on human-centric IoT and vision transformer', *J. Ambient Intell. Humanized Comput.*, 2024, doi: 10.1007/s12652-024-04797-9.

[61] X. Hu, S. Wen, and H. R. Karimi, 'An improved algorithm for deep convolutional neural network structures based on randomness', *Inf Sci*, vol. 713, 2025, doi: 10.1016/j.ins.2025.122162.

[62] Y. Tai, Y. Tan, E. Zou, B. Lei, Q. Fan, and Y. He, 'Where to model the epistemic uncertainty of Bayesian convolutional neural networks for classification', *Neurocomputing*, vol. 583, 2024, doi: 10.1016/j.neucom.2024.127568.

[63] M. Shi, X. Zeng, J. Ren, and Y. Shi, 'A multi-scale residual capsule network for hyperspectral image classification with small training samples', *Multimedia Tools Appl*, vol. 82, no. 26, pp. 40473–40501, 2023, doi: 10.1007/s11042-023-15017-5.

[64] X. Li, C. Zhao, X. Deng, and W. Jiang, 'VTFR-AT: Adversarial Training With Visual Transformation and Feature Robustness', *IEEE Trans. Emerging Topics Comp. Intell.*, vol. 8, no. 4, pp. 3129–3140, 2024, doi: 10.1109/TETCI.2024.3370004.

[65] S. Yuan, Y. Chen, C. Ye, M. W. Bhatt, M. Sardeshmukh, and M. S. Hossain, 'Cross-modal multi-label image classification modeling and recognition based on nonlinear', *Nonlenier Eng.*, vol. 12, no. 1, 2023, doi: 10.1515/nleng-2022-0194.

[66] T. K. Dutta, D. R. Nayak, and Y.-D. Zhang, 'ARM-Net: Attention-guided residual multiscale CNN for multiclass brain tumor classification using MR images', *Biomed. Signal Process. Control*, vol. 87, 2024, doi: 10.1016/j.bspc.2023.105421.

[67] L. Zhu *et al.*, 'Bayes-CAL: Robust Cross-Modal Alignment by Bayesian Approach for Few-Shot OoD Generalization', *Int J Comput Vision*, vol. 133, no. 10, pp. 7076–7109, 2025, doi: 10.1007/s11263-025-02527-y.

[68] H. Wang *et al.*, 'Optimized lightweight CA-transformer: Using transformer for fine-grained visual categorization', *Ecol. Informatics*, vol. 71, 2022, doi: 10.1016/j.ecoinf.2022.101827.

[69] Z. Jin and Y. Wei, 'UMPA: Unified multi-modal prompt with adapter for vision-language models', *Multimedia Syst*, vol. 31, no. 2, 2025, doi: 10.1007/s00530-025-01707-7.