International Journal of Formal Sciences: Current and Future Research Trends (IJFSCFRT)
ISSN: 2790-7945
https://ijfscfrtjournal.isrra.org/index.php/Formal_Sciences_Journal/index

Architecture of Deep Learning Algorithms in Image Classification: Systematic Literature Review

Vincent Mbandu Ochango[a*], John Gichuki Ndia[b]

[a,b]*Murang'a University of Technology, 75, Murang'a and 10200, Kenya*
[a]*Email: ochangovincent@gmail.com*
[b]*Email: jndia@mut.ac.ke*

Abstract

The development of deep learning algorithms has led to major improvements in image classification, a key problem in computer vision. In this study, the researcher provide an in-depth analysis of the various deep learning method architectures used for image classification. By efficiently learning hierarchical representations straight from raw image data, deep learning has brought about amazing performance gains across a wide range of applications, therefore revolutionizing the discipline. The objective was to review how different architectural choices impact the performance of deep learning models in image classification. Journals and papers published by IEEE access, ACM, Springer, Google scholar, Wiley online library, and Springer between 2013 and 2023 were analyzed. Sixty two publications were chosen based on their titles from the results of the search. The results show that more complex designs usually have better accuracy, but they may also be prone to overfitting and so benefit from regularization methods. Convolutional layers for feature extraction, pooling layers for down sampling and lowering spatial dimensions, and fully linked layers for classification are typical architectural components in deep learning algorithms for image classification. The common occurrence of skip connections in residual networks allows for a more uniform gradient flow and the training of more complex models. Models' discriminatory skills may be improved with the use of attention processes that help them zero down on important parts of a picture. In conclusion to prevent overfitting, regularization techniques like batch normalization and dropout are often used. Improved feature propagation and targeted learning, enabled by skip connections and attention techniques, greatly boosts model performance.

Keywords: Feature extraction; deep learning; deep neural network architectures; convolution neural network; and transfer learning.

Received: 9/24/2023
Accepted: 10/29/2023
Published: 11/10/2023

----

* Corresponding author.

1. Introduction

The human nervous system and brain structure serve as inspiration for the Neural Network machine learning (ML) technique. It has processing units that are separated into an input layer, a concealed layer, and an output layer. Each layer contains a set of connected nodes or units. A weight is assigned to each connection. Each node multiplies the inputs by its own weight before adding them all up. The activation function, typically a sigmoid, tan, or ReLU, is then used to do some sort of change on the total [1]. For the purpose of computing partial derivatives of the error delta with respect to particular weights, these functions are utilized since their derivative is more amenable from a mathematical perspective. Similarly, the input is compressed into a small output range or choice with the sigmoid and tanh functions (0, 1, and 1, C1, respectively). As the outputs plateau or saturate at their respective thresholds, they implement a nonlinearity with a saturated behavior. In contrast, f (x)=max(0, x) for ReLu displays both saturating and non-saturating characteristics. The result of the function is used as input for the next layer's processing unit. The output of the last layer is what is used to solve the problem [2, 3].

The convolution neural network utilizes the principle of forwarding and backward propagation which is built around the theory of multilayer perceptron or vanilla neural network. The network can as well be used in text classification, natural language processing, video analytics, and image processing. The way the brain reads and processes images is not entirely mimicked by the vanilla neural network even though it learns highly complex functions [4]. In image-related tasks, extraordinary results are achieved by the convolution neural network since it uses the working principles of the animal visual system [5]. The image processing domain has not experienced a significant breakthrough with the concept of the vanilla neural network. Mathematically, the product of 2 matrices when summed together it is known as convolution[6].Pattern recognition, grouping, dimension reduction, computer vision, NLP, regression, predictive analysis, etc. are only few of the many applications of Neural Networks.

The goal of this work is to provide a comprehensive overview of the technologies, approaches, and models currently in use for image classification based on a variety of factors [9]. This study is broken down into several sections, such as a Literature Review of related prior research, and a description of the algorithms used for image classification [10]. Finally, the study wraps up with a discussion of the strengths and weaknesses of each model for use in image classification, as well as their respective applications going forward[11].

The remainder of the paper is laid out as follows. Section 2 provides some context, Section 3 describes the techniques used, Section 4 describes the findings, Section 5 discusses the findings, and Section 6 provides a summary and suggestions for further research.

2. Related Work

Vinoth and his colleagues 2023 [12] proposed a paper in which pneumonia disease classification was done using Alexnet deep learning algorithm. Pneumonia is a dangerous illness that weakens the respiratory system. Pneumonia is thought to be caused mostly by viral germs, and doctors who treat the disease often employ radiotherapy [13]. Pneumonia diagnosis requires evaluating chest X-rays. The absence of a mobile diagnosis tool is the biggest obstacle for nations trying to slow the spread of Pneumonia. Evidence that automated diagnostic systems are needed and valued by the research community [14]. Two-thirds of patients with Pneumonia, according to a study by the World Health Organization (WHO), still need access by a radiologist. In order to solve these issues, the Alexnet model was suggested here. Early Pneumonia detection using the Alexnet model was the major objective of the suggested method. Kaggle's database, which includes both training and tested data, was used to implement the suggested solution [15]. Several different performance criteria were used in the experiments to gauge the Alexnet model's efficacy and precision. From the results collected, it can be shown that the suggested Alexnet model was quite accurate, with a 96% success rate, 93% precision, and 98% recall [16].

Image classification, image analysis, clinical archives, and beholding are just some of the many areas where deep learning can be advantageous over machine learning approaches [17]. Because digital images are so widely used as data sources in hospitals, medical photograph archives are growing rapidly. Medical images are widely used for identification and research, and digital images play an important role in predicting the severity of a patient's condition. New imaging capabilities have made automatic medical image recognition an unanswered research question in computer vision [18]. A suitable most important for sorting medical pictures into their proper categories. In recent years, Deep Learning has become increasingly popular as a means of teaching computers to recognize specific types of medical imagery [19]. Medical images of different body parts are classified using Google net and Mobile net, and the resulting models are compared. This method of image classification is helpful for making educated guesses about what category (or classes) random unfamiliar images belong to[20]. The experimental results demonstrate that the mobile net method is optimal for classifying a wide variety of medical images [21].

A deep learning framework is built for the purpose of classifying medical images, which requires the training of images. Diagnosis is currently one of the most pressing needs, and certain issues need to be investigated. As a result, a technique based on deep convolution neural networks was developed to aid medical professionals in making sound judgments. When compared to the results obtained using state-of-the-art methods, the results obtained using the mobile net strategy were superior for the same dataset [22].

In the end, the researcher compared algorithms while testing and on mobile devices. MobileNet's precision surpasses that of Google Net[23]. Cancer is a major health problem all around the globe. Breast cancer is the most common disease in women caused by the unchecked growth of abnormal cells. Identifying and categorising breast cancer are formidable challenges[24]. Therefore, several modern computational methods have been used to the detection and classification of breast cancer. These methods include k-nearest neighbour (KNN), support vector machine (SVM), multilayer perceptron (MLP), decision tree (DT), and genetic algorithms [25].

However, the precision with which each technique may be used varies. In addition, the researchers proposed a unique VGGNet-based convolutional neural network (CNN) model. Overfitting occurs in the present VGGNet-16 model due to the network's 16 layers. As a result, the researchers advocate using the VGGNet-12 model for identifying breast cancer[26]. The VGGNet-16 model suffers from the issue of overfitting the breast cancer classification dataset. The study cut down on the VGGNet-16 model's, total number of different layers to address the overfitting flaws in the original model[27]. The research presented a new version of the VGGNet model, the VGGNet- 12 model, as a response to the development of other VGGNet models, such as VGGNet-13 and VGGNet-19[28]. The model's efficacy was evaluated in comparison to the CNN and LeNet models using the breast cancer dataset. Compared to the model utilized in this work, the simulation results show that the suggested VGGNet-12 model produced an accuracy score of 0.85. Experiment results indicated that the proposed VGGNet-12 model performed well in identifying breast cancer with respect to numerous features [29].

3. Methodology

The researcher conducted a systematic literature review, a methodical and organized strategy to collecting, analyzing, and synthesizing data from previously published research on a certain subject or issue. To do this, the researcher methodically searched for studies that are relevant to the topic, evaluated their quality, extracted pertinent data, and summarized the results. The method began with review preparation, execution, and reporting. Review process details what questions will be asked and how they will be answered. Understanding how do different architectural choices, such as network depth, convolutional layers, pooling layers, and activation functions, impact the performance of deep learning models in image classification was the focus of this study. Both the databases and the search criteria were taken into account during the creation of the search method. The year and type of publication were used to set inclusion and exclusion criteria. A comprehensive search of scholarly and popular publications, such as scholarly journals and conference proceedings was conducted [37].

3.1 Research Questions

The following are some of the research questions that will be answered in this investigation;

RQ1. How do different architectural choices impact the performance of deep learning models in image classification?

RQ2. What is the process involved in image classification?

RQ3. What are the common architectural components utilized in deep learning algorithms for image classification?

3.2 Search Criteria

Massive amounts of information have already been amassed online, and much more have been created as a direct result of Google searches. IEEE Xplore Digital Library, Google Scholar, MDPI Open Access, the Wiley online library, EBSCO, the ACM digital library, Springer, and Taylor & Francis Online databases were mined for the evaluated literature [38]. When searching for information, the researcher used criteria based on how different architectural choices impact the performance of deep learning models in image classification. White papers, conference papers, journal articles, books, and book chapters were all generated. The titles, abstracts, and full texts of submitted works were evaluated for eligibility. Publications from 2013-2023 were chosen for this study

[39]. Documents retrieved, rejected, and accepted from each database are shown in Table 1.

Table 1: List of Databases.

Database	*Number of documents retrieved*	*Discarded duplicates and/or older than 2017*	*Considered for inclusion*
IEEE Explore	*109*	*103*	*15*
Google Sholar	*54*	*49*	*5*
ACM digital library	*64*	*57*	*7*
Wiley online library	*77*	*74*	*4*
Taylor & Francis	*62*	*58*	*4*
Springer	*21*	*19*	*2*
Elsevier	*79*	*68*	*11*
Snowballing			*14*
	466	*428*	*62*

3.3 Information Sources

The Google search was conducted using the criterion of how different architectural choices impact the performance of deep learning models in image classification. Since all documents containing the keyword everywhere were retrieved, the search resulted in tens of thousands of documents. The researcher just looked at the first few pages of each database. Digital objects with the same digital object identifier (DOI) were treated as duplicates and deleted [40]. All papers published until 2023 were included in the search. When feasible, the search phrases have been looked for in the article's title, abstract, and body. Before narrowing down the list, the researcher also did a snowball search on some of the cited sources in the chosen publications [41].

3.4 Inclusion and Exclusion Criteria

The review covers publications published between 2013 and 2023 and is limited to those that include the search keyword anywhere in the text. After reading the titles, abstracts, introductions, and entire articles, the recovered papers were sorted into relevant and non-relevant piles. Relevance was used to choose which documents to include. Unpublished research on how different architectural choices impact the performance of deep learning models in image classification, materials from other domains, and documents published in other databases were not included in the review. Documents mentioned in the appropriate literature were also collected, a process known as snowballing. Vertical search, in which citations serve as a roadmap for learning more about a topic, is known as snowballing [42].

4. Results and Discussion

The Results section of a systematic literature review provides a synopsis and analysis of the results collected from the studies that were included in the review. This part is crucial because it gives readers an unbiased and all-encompassing summary of the information around a particular research question or subject.

4.1 RQ1: How do different architectural choices impact the performance of deep learning models in image classification?

Convolutional Neural Network are often used in deep learning algorithms for image categorization because of their impressive performance in this area. Convolutional Neural Networks are artificial neural networks that are programmed to automatically learn hierarchical representations of visual input. Convolutional layers, pooling layers, and fully linked layers are all part of a standard CNN's architecture when it comes to image categorization. Table 2 summarizes the main architectural decisions and their broad effects on image classification models, with explanatory text. The explanations offered show the complexity and interdependence of various decisions, while the tabular style seeks to encapsulate the implications of architectural choices. The best architecture to use in reality is conditional on the nature of the issue at hand, the size of the dataset, and the resources at hand. Model complexity, training efficiency, and generalization capacity are three factors that are commonly combined in a well-designed architecture.

Table 2: Architectural Choice.

Architectural Choice	Impact on Performance	Explanations	N0. Of Papers	References
Network Depth	Deeper networks can learn more complex features but may suffer from vanishing/exploding gradients and require careful initialization and regularization techniques.	Increasing the depth of a neural network allows it to learn intricate patterns and features from images. However, very deep networks might encounter issues with gradient propagation during backpropagation, which can lead to training difficulties. Proper initialization methods and regularization techniques are necessary to mitigate these problems.	10	[1, 2, 4, 8, 9] [11, 16]
Convolutional Layers	Convolutional layers are essential for capturing spatial hierarchies and local patterns in images.	Convolutional layers perform localized feature extraction by applying filters to different parts of the image. They excel at detecting edges, textures, and other spatially relevant information, making them crucial for image classification tasks.	8	[11, 21, 34] [58, 19, 11, 26]
Pooling Layers	Pooling layers reduce spatial dimensions and enhance translation invariance, enabling the network to focus on important features.	Pooling layers down sample the feature maps, reducing the computational load and providing a degree of translation invariance, which makes the model more robust to variations in object position within the	5	[21, 31, 4, 28, 15, 51, 46]

		image.		
Activation Functions	Non-linear activation functions introduce complexity and enable the model to learn complex mappings from inputs to outputs.	Activation functions like ReLU introduce non-linearity to the network, enabling it to capture intricate relationships between features. Without non-linear activations, the network would be limited to learning linear transformations.	9	[27, 3, 4, 28, 5, 1, 56]
Batch Normalization	Batch normalization accelerates training by reducing internal covariate shift and stabilizing gradient flow.	Batch normalization normalizes the activations of each layer, which helps in faster convergence during training by reducing the internal covariate shift. It also has a regularizing effect that can lead to improved generalization.	15	[23, 53, 44, 58]
Skip Connections	Skip connections aid in training deep networks by alleviating vanishing gradient issues and allowing the network to learn both low-level and high-level features.	Skip connections enable gradients to bypass several layers, addressing the vanishing gradient problem in deep networks. They also facilitate the learning of lower-level features and their combination into higher-level abstractions.	3	[52, 55, 60] [47, 35, 59, 56, 33, 49]
Data Augmentation	Data augmentation increases model robustness and generalization by exposing the model to various data variations.	Data augmentation involves applying transformations to the training data (e.g., rotations, flips, crops) to artificially increase the diversity of the training set. This helps the model become more invariant to different transformations that might be encountered during inference.	2	[2, 25, 51, 41, 33, 39, 36, 33, 44]

Dropout	Dropout is a regularization technique that prevents overfitting by randomly dropping units during training, forcing the network to learn more robust features.	Dropout helps prevent the model from relying too heavily on specific units, leading to better generalization. It simulates an ensemble of networks during training, which can improve the model's ability to handle various inputs.	2	[17, 13, 10, 20, 18, 34, 30, 41, 3]
Learning Rate Schedule	Learning rate schedules adjust the learning rate over time, influencing the convergence speed and stability of the training process.	Learning rate schedules start with a relatively high learning rate and then decrease it gradually during training. This helps to ensure quicker convergence in the early stages while allowing for more precise updates as training progresses.	3	[12, 5, 6, 41, 32, 42, 51, 40, 52]
Optimizer Choice	Optimizers like Adam, SGD, and RMSProp affect how the model's weights are updated during training, impacting convergence speed and stability.	Different optimizers have distinct algorithms for updating the model's weights based on gradients. The choice of optimizer can influence the speed of convergence and the stability of the training process.	2	[1, 11, 12, 29, 53, 48, 28, 17, 37]
Initial Weighting	Proper weight initialization methods contribute to smoother training by aiding gradient propagation and convergence.	Initializing weights with appropriate methods (e.g., Xavier, He initialization) ensures that gradients flow consistently through the network and helps the model converge faster. Poor initialization can lead to slow convergence or even prevent learning altogether.	3	[7, 14, 54, 50, 31, 19, 26, 55, 43]
			62	

The needs of a given image classification task will dictate how the architecture has to be modified and expanded. To further enhance the model's performance, more sophisticated methods including batch normalization, residual connections, and attention processes may be implemented [50].

The effectiveness of deep learning models for image categorization may be drastically altered by tweaking various architectural parameters. The model's accuracy improves when the network depth is increased, since then it can learn more complicated information. However, very deep networks risk overfitting on smaller datasets and experiencing vanishing gradients. Capturing local features and textures relies heavily on convolutional layers, with more layers resulting in superior feature extraction [51]. However, the training computational cost also grows with the number of convolutional layers. Down sampling the feature maps is the job of pooling layers, which also reduces computation and expands the receptive field. By selecting the appropriate pooling methods, we can save the most important data while deleting the less important geographical features [52].

Training convergence and dealing with the vanishing gradient issue are impacted by the choice of activation functions since they add non-linearity into the model. The network's efficiency is drastically altered by popular options like ReLU and its variations. By standardizing activations, normalization layers like batch normalization help to stabilize and speed up training, resulting in better generalization and quicker convergence [53]. ResNet-like topologies, which have skip connections, allow for extremely deep network training and help prevent disappearing gradients. Network width, here meaning the total number of channels or neurons, affects the model's ability to learn several characteristics at once but at the expense of increased computing complexity [54].

Furthermore, when the target dataset is tiny, it is possible to dramatically enhance performance by using pretraining on a big dataset and fine-tuning on the target dataset. Overfitting may be avoided and generalization performance can be improved by using regularization methods including dropout, weight decay, and data augmentation. Convergence time and performance are affected by the optimizer and learning rate schedule that is used. Finding the right mix is the key to effective training. Object categorization in complicated settings is improved by the model's ability to zero in on relevant parts of the picture using attention processes [55].

The success of a deep learning model for image categorization ultimately rests on the careful balancing of architectural options that are specific to the dataset and the job at hand. The best outcomes can only be attained via experimentation and fine-tuning [56].

4.2 RQ1: What is the process involved in image classification?

The process of image classification consists of many steps: acquisition of images, segmentation of those images, feature extraction, and ultimately classification. The steps required for Image Classification are shown in Figure 1.

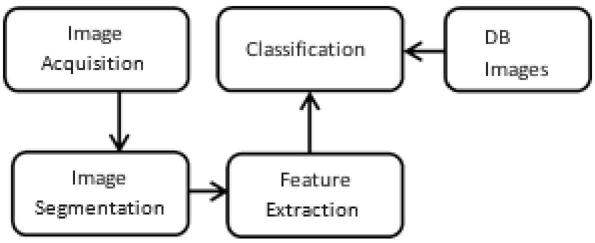

Figure 1: Process Involved in Image Classification.

I. Image Acquisition

In Image Acquisition, digital equipment like cameras and scanners are used to capture and collect pictures. These photos are collected and stored in a database for later use in the classification process; a larger database yields more precise classifications.

II.Image Segmentation

Image Segmentation simplifies and streamlines the analysis of images. This method involves breaking up a picture into smaller pieces, or image segments that may then be handled separately. Otsu's Algorithm, k-means clustering, etc., are only a few of the numerous approaches that may be used to complete the Image Segmentation. Clusters of comparable objects are created using K-means clustering. This technique for grouping data into clusters uses measures of similarity between its components to do so. The first phase of the k-means clustering technique is to choose the k value, the second step is to initialize centroids, and the third step is to choose the data group and locate the mean.

III.Feature Extraction

The term feature extraction refers to the wide variety of techniques used to extract features from data. Following picture Segmentation, the input picture is passed on to the feature extraction module, where relevant features and information are retrieved. Extracting features such as texture, color, leaf shape, edges, etc. is feasible.

IV.Classification

After feature extraction, the classifier will sort the characteristics based on the image class. Classifiers like the Convolutional Neural Network (CNN), the Support Vector Machine (SVM), the Artificial Neural Network (ANN), the Naive Bayes classifier, the K-Nearest Neighbor (KNN), the Decision Tree classifier, etc. are all discussed below.

a) Convolution Neural Network

CNNs have several uses in the fields of Image Analysis and Computer Vision. They are used for Image Analysis, a process that involves analyzing and identifying various aspects inside pictures. Medical image analysis, video identification and analysis, and computer vision are just a few of the many possible uses. Due to its superior accuracy, CNN is often employed for this purpose. The Input Layer, Convolution Layer, Pooling Layer, Fully Connected Layer, and Output Layer are the building blocks of the CNN model [57].

i. Convolution Layer

In order to extract features from the input pictures/training images, the Convolution Layer is employed. Each convolution layer will include filters for feature extraction from pictures. Some operations are carried out on the input picture and a filter of size NxN in the Convolution Layer. The dot product is calculated between the regions of the input picture and the filter as the filter is slid over the image. The result of this procedure is a feature map, a representation of the picture data. Several elements of the picture are derived from the

Convolution Layer [58].

ii. Pooling Layer

By sliding the filter over each channel of the feature maps and summarizing the features within the filter region, we form a pooling layer, which is used to reduce the size of the feature map in order to reduce computational costs after the Convolutional layer. By lowering the image's spatial size using a pooling layer, processing time may be saved. Max pooling, one of several kinds of pooling layers (others include sum pooling, average pooling, and so on), is used to determine the biggest element in the feature map. The sum pooling algorithm adds together the values in a specified range. When items in a certain region are pooled together, their average is calculated [59].

iii. Fully Connected Layer

The last few layers of a Convolutional Neural Network Architecture are the Fully Connected Layer and the Output Layer. The pooling layer's output is the input of the fully connected layer. After the matrix has been flattened, it passes via Fully Connected layers, which is where the actual calculations happen. This is the classification procedure. The purpose of this layer is to label the pictures [60].

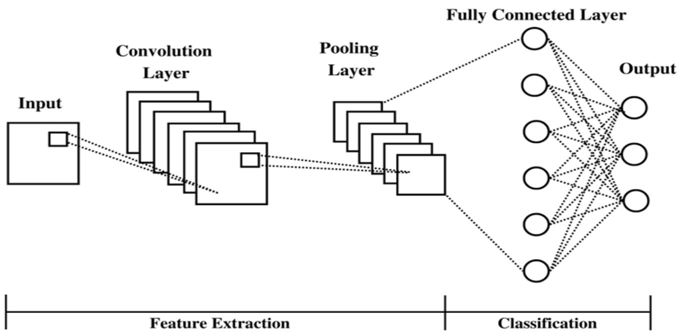

Figure 2: The Architecture of the Convolution Neural Network.

b) Support Vector Machine (SVM)

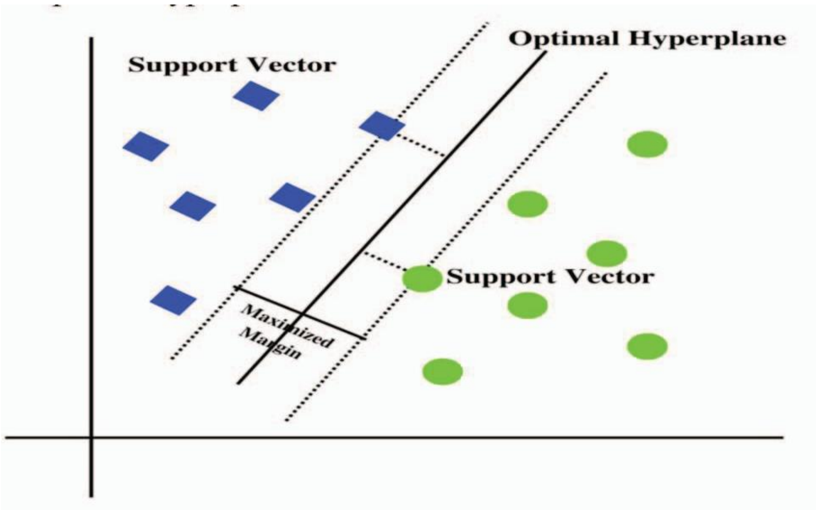

Figure 3: The Support Vector Machine.

The SVM technique generates a decision boundary that partitions the n-dimensional space into classes, allowing us to properly classify the latest data point. The hyperplane is the decision boundary. When generating a hyperplane, the Support Vector Machine decides which extreme vectors to utilize. Figure 2 depicts a Support Vector Machine, which uses a hyperplane to classify data into two groups. The image classification, face recognition, and other applications of the Support Vector Machine method. Linear SVMs and non-linear SVMs are the two main varieties of Support Vector Machines. Non-Linear SVM is used when a dataset cannot be divided into two categories by a straight line, whereas Linear SVM is used whenever this division is possible.

The vectors or data points that are closer to the hyperplane and impact the hyperplane location are termed Support Vectors, and the optimal decision boundary in categorizing the data is called a hyperplane [61].

Red and green are the labels for a dataset with two features, x1 and x2. This calls for a Classifier that assigns a value of Red or Green to the new data point. The SVM Algorithm may be used to locate the optimal decision boundary, also known as a hyperplane.

The vectors that are closest to the hyperplane are referred to as support vectors. Margin refers to the space between the support vectors and the hyperplane. An ideal hyperplane is one in which the margin is maximized [62].

The standard procedure for image classification is laid forth in table 3 provided below. The researcher then provides paragraph-length explanations of each stage of the method that follow the table.

Table 3: Image Classification Process.

Step	Description	No. of papers	References
1. Data Collection	Gather a diverse set of labeled images for training.	10	[1,5,16,19,26,27, 61,62,63,64]
2. Data Preprocessing	Resize, normalize, and augment images for consistency.	12	[65,38,39,40,41,42,55, 68,69,70,71,43]
3. Feature Extraction	Extract relevant features using techniques like CNNs.	7	[3,44,45,46,47, 48,31,23]
4. Model Selection	Choose an appropriate classification algorithm/model.	5	[8,25,56,57,14]
5. Model Training	Train the model on the labeled images using the features.	10	[4,10,12,20,21,28, 29,30,31,32]
6. Model Evaluation	Assess model performance using validation data.	10	[34,35,36,37,46, 50,52,55,60,66,67]
7.Hyperparameter Tuning	Fine-tune model parameters for better performance.	3	[33,3,44]
8. Prediction	Apply the trained model to classify new, unseen images.	4	[45,46,47,48]
9. Post-processing	Refine predictions or apply additional filters.	1	[31]
		62	

Classifying pictures into categories or groups that have already been established is a basic problem in computer vision known as image classification. The first step is data collection, which entails gathering and labeling a large variety of photos that are representative of each category. Data preprocessing is applied to this annotated dataset, which includes things like scaling photos to the same dimensions, standardizing pixel values, and maybe enriching the data via methods like rotation and flipping.

Next is feature extraction, which entails mechanically removing important aspects of a picture. In most cases, this is accomplished by using a Convolutional Neural Network (CNN). Following feature extraction, an appropriate classification technique or model is chosen. This may be a deep learning model like a Convolutional Neural Network (CNN) or a more conventional machine learning technique like a Support Vector Machine (SVM).

In order for a model to understand the underlying patterns and connections between features and classes, it must be fed the labeled pictures and their extracted features. After training a model, it is tested using validation data to see how well it performs. To improve performance on the validation set, the next step is hyperparameter tweaking.

The improved model may then be used to make predictions on as-yet unseen photos. Before being categorised by the trained model, these photos undergo the identical preprocessing and feature extraction processes. Finally, post-processing techniques may be performed to the findings to improve the accuracy of the model's predictions or to apply additional filters.

In conclusion, in order to properly classify photos into preset categories, a series of stages from data collection through post-processing are required for image classification. Each stage is essential to producing reliable categorization outcomes.

4.3 RQ3: What are the common architectural components utilized in deep learning algorithms for image classification?

Several shared architectural components are used by deep learning algorithms for image classification to efficiently analyze and decode visual data. Using convolutional layers is a crucial part of the framework. Learnable convolutional kernels allow Convolutional Neural Networks (CNNs) to automatically discover local patterns and features inside a picture. These kernels "slide" across the input picture, using convolutions to pull out useful characteristics.

Layers that collect water are also important. Through down sampling, the spatial dimensions and computational complexity of the feature maps extracted from the convolutional layers are decreased.

In order for the network to properly capture more abstract aspects, pooling helps the model concentrate on the most important information while eliminating the less significant spatial details.

Models may learn intricate correlations between features when non-linearity is introduced through activation functions. Rectified Linear Units (ReLU), Leaky ReLU, and variations are popular activation functions because they have been found to be effective in picture classification tasks.

Table 4: Architectural Components.

Component	Description	No. of papers	References
Convolutional Layers	These layers apply convolutional filters to the input image, capturing spatial hierarchies and detecting features like edges, textures, and patterns.	11	[1, 5, 16, 19, 26, 27, 61, 62, 33, 24, 55]
Pooling Layers	Pooling layers (e.g., MaxPooling, AveragePooling) reduce the spatial dimensions of the feature maps, retaining essential information and reducing computation.	14	[4,10,12,20,21,28, 34,35,36,37,46, 55,60,36,37]
Fully Connected Layers	These layers are traditional neural network layers where each neuron is connected to every neuron in the previous and following layers.	11	[38,39,40,41,42,55, 68,69,10,21,43]

Activation Functions	Activation functions (e.g., ReLU, Sigmoid, Tanh) introduce non-linearity, enabling the network to learn complex relationships within the data.	8	[3,44,45,46,47, 48,31,13]
Batch Normalization	Normalizes the activations of each layer, improving convergence and training stability, often accelerating training and enhancing generalization.	7	[49,50,51,52,53,54, 55]
Dropout	Dropout randomly deactivates a portion of neurons during training, reducing overfitting by forcing the network to learn more robust features.	5	[8,25,56,57,14]
Skip Connections	Also known as residual connections, these allow gradients to flow more directly, aiding in the training of very deep networks like ResNet.	5	[29,30,31,32,33]
Global Average Pooling	Replaces fully connected layers with a global average pooling layer, reducing the spatial dimensions and aiding in better feature extraction.	2	[50, 52]
		62	

The activations in each layer may be standardized with the use of normalization layers, such as batch normalization or layer normalization. As a result, training is more stable and progresses more quickly, enhancing convergence and generalization. By eliminating the internal covariate shift, normalization improves the model's ability to learn.

It has been shown that skip connections help deep architectures. The gradient flow may be expedited by cutting over unnecessary layers thanks to these shortcuts. This prevents training from being halted due to vanishing gradients and allows for very deep networks, which improves the collection of subtle patterns and features.

Furthermore, fully linked layers at the end of the network are often used by deep learning algorithms for picture classification. High-level information learnt by the convolutional layers are combined in these layers, and predictions for picture classes are made.

The use of attention processes in image classification frameworks has grown in recent years. By directing the model's attention to parts of the picture that are more relevant to the classification job, attention mechanisms improve the model's discriminative skills.

Finally, regularization methods like dropout and data augmentation are used in many state-of-the-art picture classification models. By inactivating neurons at random during training, dropout reduces overfitting, while data augmentation artificially expands the training dataset by applying random changes to the pictures, leading to a more robust and generalizable model. Deep learning algorithms are able to successfully learn from vast volumes of visual data and reach state-of-the-art performance on picture classification tasks because they combine these architectural components. Table 4 outlines common architectural components utilized in deep learning algorithms for image classification In order to classify images, deep learning algorithms often use the following common architectural components, as shown in table 4. Modern convolutional neural networks (CNNs) rely on these building blocks to reliably extract informative information from pictures and make precise predictions. Convolutional layers filter the input picture to detect fine details like edges and textures.

Using pooling layers, the spatial dimensions of the feature maps may be decreased while still keeping the necessary information. Traditional neural network components, such as linking neurons across layers, are introduced in a fully connected layer, allowing for the processing of more abstract representations. Non-linearity introduced by activation functions allows the model to understand intricate connections within the data. By standardizing layer activations, batch normalization guarantees consistent and effective training. Dropout is a method for preventing overfitting in which neurons are inactivated at random during training. The training of deep networks is made possible via skip connections, also known as residual connections, which improve the propagation of gradients. In order to improve feature extraction, global average pooling may be used in lieu of completely linked layers. Together, these parts construct robust picture classification models, able to learn hierarchical representations and provide reliable predictions. This combination has been crucial to producing state-of-the-art results on a wide range of picture datasets, and it has also revolutionized image categorization.

4.4 Discussion

The discussion section is an essential part of a systematic literature review since it is where the results from the included research are interpreted and synthesized. This section of the research report is crucial since it is where the writers draw their own conclusions based on the studies they analyzed. Discussing the study questions or goals, contrasting and contrasting the results, identifying patterns and trends, and providing explanations for any discrepancies or contradictions among the studies are all possible in the discussion part.

i. Architectural Choices

The effectiveness of deep learning models for image categorization is very sensitive to the architecture that is used. The model may be trained on more complicated and abstract characteristics if the network depth is increased by adding additional layers. However, problems like overfitting on short datasets and vanishing gradients may arise with highly deep networks. To identify small-scale patterns in pictures, convolutional layers are essential. The model's feature-capture performance may improve with more convolutional layers, but this comes at the cost of increased computational complexity. The spatial dimensions of feature maps are reduced by using pooling layers, which helps to maintain vital information while decreasing computational cost. It is critical to strike a balance between feature retention and downsampled data when using a pooling strategy. Activation functions allow the model to be non-linear, which is necessary for modeling dynamics. Training convergence and the model's robustness in the face of vanishing gradients may be affected by activation functions like ReLU and its derivatives.

ii. Image Classification Process

There are numerous crucial stages in picture categorization. Images are gathered and categorized with categories and classes in order to create the dataset. Separate training and validation sets are constructed from the labeled data. Then, taking into account the requirements in RQ3, the necessary architectural components are developed for a deep learning model, most often a convolutional neural network (CNN). Random weights are used to kick off the training process once the model has been setup. The loss function is the difference between the predicted and true labels, and during training the model learns to improve its parameters to reduce this difference. The optimization is carried out with the help of a tailored optimizer (e.g., Adam, SGD) and learning rate schedule. Overfitting may be avoided with the use of regularization strategies like dropout and weight decay. To enhance its performance on the training data, the model repeatedly adjusts its weights using backpropagation and gradient descent. Following training, the model's generalization capabilities and overall performance are assessed using a separate validation or test set.

iii. Deep Learning Architectural Components

For image categorization, deep learning systems use a few standard building blocks. In order to extract features from an input picture, convolutional layers must first slide learnable convolutional kernels over it. The geographical dimensions and computational complexity of the feature maps are downsampled using pooling layers, but the key information is preserved. Non-linearity is introduced to the model through activation functions like ReLU and its derivatives, which facilitates the learning of complicated patterns. By standardizing activations across layers, normalization layers like batch normalization assist stabilize and speed up the training process. In order to prevent the occurrence of the dreaded vanishing gradient when training very deep networks, skip connections are used. High-level characteristics are combined in fully linked layers at the conclusion of the network. The model's discriminatory skills may be improved by the use of attention mechanisms that enable it to zero down on certain areas of a picture. Dropout and data augmentation are two regularization methods that may be used to avoid model overfitting and increase generalization accuracy. Deep learning algorithms are able to efficiently learn from visual data and achieve state-of-the-art performance in image classification tasks because they combine these architectural components.

5. Conclusion

Significant developments and improvements in the area of image classification are revealed by an analysis of the architecture of deep learning algorithms. Due to its capacity to learn complex hierarchical features from pictures and attain state-of-the-art performance on a wide variety of datasets, Convolutional Neural Networks (CNNs) have quickly become the dominant architecture. In addition, researchers have employed transfer learning with pre-trained models such as VGG, ResNet, and EfficientNet to construct accurate models with minimal computing effort. By merging many models, ensemble approaches like model averaging and boosting show potential in further enhancing classification accuracy. As a result of their use in cleaning up messy and complicated datasets, attention mechanisms have also risen in favor. The computing requirements of deep learning are discussed, with an emphasis on the need of efficient architectures and model compression approaches. By adding alterations to the training data including flipping, rotation, and color tweaks, data augmentation approaches play a crucial role in avoiding overfitting and improving model generalization. Image classification has come a long way, but there is still work to be done in areas like managing massive datasets, addressing class imbalance, and guaranteeing the interpretability of deep learning models. The conclusions of this study suggest that researchers in this field should devote their attention toward finding solutions to these problems. If we want deep learning algorithms to be useful and widely used, we need to keep working to increase model efficiency and decrease their dependence on computing resources. In sum, the paper demonstrates how deep learning algorithms have progressed and where the area of image categorization may go in the future.

References

[1] M. A. Arefeen, S. T. Nimi, M. Y. S. Uddin, and Z. Li, "A Lightweight Relu-Based Feature Fusion For Aerial Scene Classification," in *2021 IEEE International Conference on Image Processing (ICIP)*, Anchorage, AK, USA: IEEE, Sep. 2021, pp. 3857–3861. doi: 10.1109/ICIP42928.2021.9506524.

[2] M. Kaloev and G. Krastev, "Comparative Analysis of Activation Functions Used in the Hidden Layers of Deep Neural Networks," in *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Ankara, Turkey: IEEE, Jun. 2021, pp. 1–5. doi: 10.1109/HORA52670.2021.9461312.

[3] Y. Sun, "The role of activation function in image classification," in *2021 International Conference on Communications, Information System and Computer Engineering (CISCE)*, Beijing, China: IEEE, May 2021, pp. 275–278. doi: 10.1109/CISCE52179.2021.9445868.

[4] Y. Tang, L. Tian, Y. Liu, Y. Wen, K. Kang, and X. Zhao, "Design and implementation of improved CNN activation function," in *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*, Changchun, China: IEEE, May 2022, pp. 1166–1170. doi: 10.1109/CVIDLICCEA56201.2022.9824061.

[5] A. Bolar, R. N. Kanuri, S. Shrihari, S. Natarajan, and K. Nagajothi, "Classification of Urban Data using Satellite Imaging," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore: IEEE, Sep. 2018, pp. 1843–1847. doi: 10.1109/ICACCI.2018.8554734.

[6] J. Feng, G. Bai, Z. Gao, X. Zhang, and X. Tang, "Automatic Design Recurrent Neural Network for Hyperspectral Image Classification," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium: IEEE, Jul. 2021, pp. 2234–2237. doi: 10.1109/IGARSS47720.2021.9554089.

[7] J. Butdee, W. Kongprawechnon, H. Nakahara, N. Chayopitak, C. Kingkan, and R. Pupadubsin, "Pattern

Recognition of Partial Discharge Faults Using Convolutional Neural Network (CNN)," in *2023 8th International Conference on Control and Robotics Engineering (ICCRE)*, Niigata, Japan: IEEE, Apr. 2023, pp. 61–66. doi: 10.1109/ICCRE57112.2023.10155616.

[8] C. Lai, T. Liu, R. Mei, H. Wang, and S. Hu, "The Cloud Images Classification Based on Convolutional Neural Network," in *2019 International Conference on Meteorology Observations (ICMO)*, Chengdu, China: IEEE, Dec. 2019, pp. 1–4. doi: 10.1109/ICMO49322.2019.9026121.

[9] Y. Harshitha, B. L. N. P. Kumar, M. V. Aparna, and A. V. Kishore, "Remote Sensing Image Classification with a Few Labeled Samples," in *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India: IEEE, Feb. 2023, pp. 1557–1561. doi: 10.1109/ICAIS56108.2023.10073806.

[10] O. Adigun and B. Kosko, "Deeper Neural Networks with Non-Vanishing Logistic Hidden Units: NoVa vs. ReLU Neurons," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Pasadena, CA, USA: IEEE, Dec. 2021, pp. 1407–1412. doi: 10.1109/ICMLA52953.2021.00227.

[11] P. Roy, S. Ghosh, and U. Pal, "A CNN Based Framework for Unistroke Numeral Recognition in Air-Writing," in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Niagara Falls, NY, USA: IEEE, Aug. 2018, pp. 404–409. doi: 10.1109/ICFHR-2018.2018.00077.

[12] R. Vinoth, S. Subalakshmi, and S. Thamaraichandra, "Pneumonia Detection from Chest X-Ray using AlexNet Image Classification Technique," in *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India: IEEE, May 2023, pp. 1307–1312. doi: 10.1109/ICICCS56967.2023.10142405.

[13] M. Feng, D. Chen, and X. Wang, "Classification of Fabric Patterns Image Based on Improved Log-AlexNet," in *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Beijing, China: IEEE, Oct. 2018, pp. 1–6. doi: 10.1109/CISP-BMEI.2018.8633260.

[14] Y. Li, W. Wei, J. Zhang, L. Zhang, and Y. Zhang, "Data-Specific Activation Function Learning for Hyperspectral Image Classification," in *2020 8th International Conference on Orange Technology (ICOT)*, Daegu, Korea (South): IEEE, Dec. 2020, pp. 1–4. doi: 10.1109/ICOT51877.2020.9468782.

[15] H. Zhao, D. Liu, and H. Li, "Efficient Integer-Arithmetic-Only Convolutional Networks with Bounded ReLU," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, Daegu, Korea: IEEE, May 2021, pp. 1–5. doi: 10.1109/ISCAS51556.2021.9401448.

[16] C. Xu, "Applying MLP and CNN on Handwriting Images for Image Classification Task," in *2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering*

(AEMCSE), Wuhan, China: IEEE, Apr. 2022, pp. 830–835. doi: 10.1109/AEMCSE55572.2022.00167.

[17] A. Stanojevic, E. Eleftheriou, G. Cherubini, S. Wozniak, A. Pantazi, and W. Gerstner, "Approximating Relu Networks by Single-Spike Computation," in *2022 IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France: IEEE, Oct. 2022, pp. 1901–1905. doi: 10.1109/ICIP46576.2022.9897692.

[18] W. Li and R. Hua, "Image classification method based on improved deep residual networks," in *2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, Shanghai, China: IEEE, Nov. 2019, pp. 380–384. doi: 10.1109/ICIIBMS46890.2019.8991495.

[19] R. Kanabarkar and M. Anandhalli, "Performance Analysis of Convolutional Neural Network for Image Classification," in *2022 International Conference on Futuristic Technologies (INCOFT)*, Belgaum, India: IEEE, Nov. 2022, pp. 1–4. doi: 10.1109/INCOFT55651.2022.10094394.

[20] M. Elsaadouny, J. Barowski, and I. Rolfes, "The Subsurface Objects Classification using a Convolutional Neural Network," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada: IEEE, Oct. 2019, pp. 0874–0877. doi: 10.1109/IEMCON.2019.8936250.

[21] S. K. P. N, D. P, A. G, Saritha, S. K, and D. R. U, "Automated matching of pixel of interest between two digital images from two different microscope imaging devices," in *2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Thiruvananthapuram, India: IEEE, Dec. 2020, pp. 96–100. doi: 10.1109/RAICS51191.2020.9332522.

[22] T. Takase *et al.*, "An Estimation Method of Road Surface Condition on Winter Expressway via Mobile Nets using In-vehicle Camera Images," in *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, Kyoto, Japan: IEEE, Oct. 2021, pp. 109–110. doi: 10.1109/GCCE53005.2021.9621866.

[23] V. A. Krishna, D. Nagajyothi, A. A. Reddy, and D. Aravind, "Image Data Classification Using Mobile Net for Big Data Analytics," in *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Ballari, India: IEEE, Apr. 2022, pp. 1–6. doi: 10.1109/ICDCECE53908.2022.9792904.

[24] M. N. Qureshi and M. Sarosh Umar, "SaRa: A Novel Activation Function with Application to Melanoma Image Classification," in *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India: IEEE, Dec. 2022, pp. 854–858. doi: 10.1109/ICACRS55517.2022.10029161.

[25] S. Zhong, C.-I. Chang, and Y. Zhang, "Iterative Support Vector Machine for Hyperspectral Image Classification," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens:

IEEE, Oct. 2018, pp. 3309–3312. doi: 10.1109/ICIP.2018.8451145.

[26] Z. He and L. Xiong, "Image Classification of Zinc Dross Based on Improved MobileNetV2," in *2021 China Automation Congress (CAC)*, Beijing, China: IEEE, Oct. 2021, pp. 1503–1508. doi: 10.1109/CAC53003.2021.9728248.

[27] Y. Zhang, Q. Hua, D. Xu, H. Li, Y. Bu, and P. Zhao, "A Complex-Valued CNN for Different Activation Functions in Polarsar Image Classification," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan: IEEE, Jul. 2019, pp. 10023–10026. doi: 10.1109/IGARSS.2019.8898534.

[28] W. Guoxin, L. Xiuli, J. Zhanglei, and H. Ruxiang, "Dongba classical ancient books image classification method based on ReN-Softplus convolution residual neural network," in *2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, Changsha, China: IEEE, Nov. 2019, pp. 398–404. doi: 10.1109/ICEMI46757.2019.9101450.

[29] A. Khan, A. Khan, M. Ullah, M. M. Alam, J. I. Bangash, and M. M. Suud, "A computational classification method of breast cancer images using the VGGNet model," *Front. Comput. Neurosci.*, vol. 16, p. 1001803, Nov. 2022, doi: 10.3389/fncom.2022.1001803.

[30] J. Shijie, W. Ping, J. Peiyi, and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," in *2017 Chinese Automation Congress (CAC)*, Jinan: IEEE, Oct. 2017, pp. 4165–4170. doi: 10.1109/CAC.2017.8243510.

[31] S. K. Roy, M. E. Paoletti, J. M. Haut, E. M. T. Hendrix, and A. Plaza, "A New Max-Min Convolutional Network for Hyperspectral Image Classification," in *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Amsterdam, Netherlands: IEEE, Mar. 2021, pp. 1–5. doi: 10.1109/WHISPERS52202.2021.9483983.

[32] X. Jia *et al.*, "Automatic Polyp Recognition in Colonoscopy Images Using Deep Learning and Two-Stage Pyramidal Feature Prediction," *IEEE Trans. Autom. Sci. Eng.*, pp. 1–15, 2020, doi: 10.1109/TASE.2020.2964827.

[33] H. H. Vo and A. Verma, "New Deep Neural Nets for Fine-Grained Diabetic Retinopathy Recognition on Hybrid Color Space," in *2016 IEEE International Symposium on Multimedia (ISM)*, San Jose, CA, USA: IEEE, Dec. 2016, pp. 209–215. doi: 10.1109/ISM.2016.0049.

[34] G. Yang, U. B. Gewali, E. Ientilucci, M. Gartley, and S. T. Monteiro, "Dual-Channel Densenet for Hyperspectral Image Classification," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia: IEEE, Jul. 2018, pp. 2595–2598. doi: 10.1109/IGARSS.2018.8517520.

[35] L. Ou and K. Zhu, “Identification Algorithm of Diseased Leaves based on MobileNet Model,” in *2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE)*, Shenzhen, China: IEEE, May 2022, pp. 318–321. doi: 10.1109/CISCE55963.2022.9851086.

[36] H. Yin, C. Yang, and J. Lu, “Research on Remote Sensing Image Classification Algorithm Based on EfficientNet,” in *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Xi’an, China: IEEE, Apr. 2022, pp. 1757–1761. doi: 10.1109/ICSP54964.2022.9778437.

[37] U. K. S, P. J, and K. S. R. S, “Systematic Review of Deep Learning Models for Dental Images,” in *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India: IEEE, Feb. 2023, pp. 287–291. doi: 10.1109/ICCMC56507.2023.10083729.

[38] M. Yasir, W. Jianhua, L. Shanwei, H. Sheng, X. Mingming, and M. Hossain, “Coupling of deep learning and remote sensing: a comprehensive systematic literature review,” *Int. J. Remote Sens.*, vol. 44, no. 1, pp. 157–193, Jan. 2023, doi: 10.1080/01431161.2022.2161856.

[39] P. V. Torres-Carrion, C. S. Gonzalez-Gonzalez, S. Aciar, and G. Rodriguez-Morales, “Methodology for systematic literature review applied to engineering and education,” in *2018 IEEE Global Engineering Education Conference (EDUCON)*, Tenerife: IEEE, Apr. 2018, pp. 1364–1373. doi: 10.1109/EDUCON.2018.8363388.

[40] I. M. Black, M. Richmond, and A. Kolios, “Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management,” *Int. J. Sustain. Energy*, vol. 40, no. 10, pp. 923–946, Nov. 2021, doi: 10.1080/14786451.2021.1890736.

[41] L. Stroebel, M. Llewellyn, T. Hartley, T. S. Ip, and M. Ahmed, “A systematic literature review on the effectiveness of deepfake detection techniques,” *J. Cyber Secur. Technol.*, vol. 7, no. 2, pp. 83–113, Apr. 2023, doi: 10.1080/23742917.2023.2192888.

[42] X. Zhai, Y. Yin, J. W. Pellegrino, K. C. Haudek, and L. Shi, “Applying machine learning in science assessment: a systematic review,” *Stud. Sci. Educ.*, vol. 56, no. 1, pp. 111–151, Jan. 2020, doi: 10.1080/03057267.2020.1735757.

[43] Y. Liu, G. Luo, and F. Dong, “Convolutional Network Model using Hierarchical Prediction and its Application in Clothing Image Classification,” in *2019 3rd International Conference on Data Science and Business Analytics (ICDSBA)*, Istanbul, Turkey: IEEE, Oct. 2019, pp. 157–160. doi: 10.1109/ICDSBA48748.2019.00041.

[44] A. S., S. A, and G. K, “Classification of Agricultural Leaf Images using Hybrid Combination of Activation Functions,” in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India: IEEE, May 2021, pp. 785–791. doi: 10.1109/ICICCS51141.2021.9432221.

[45] F. M. J. Mehedi Shamrat, Md. A. Jubair, Md. M. Billah, S. Chakraborty, Md. Alauddin, and R. Ranjan, "A Deep Learning Approach for Face Detection using Max Pooling," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India: IEEE, Jun. 2021, pp. 760–764. doi: 10.1109/ICOEI51242.2021.9452896.

[46] Y. H. Tsai, N. Y. Lyu, S. Y. Jung, K. H. Chang, J. Y. Chang, and C. T. Sun, "Deep Learning Based AOI System with Equivalent Convolutional Layers Transformed from Fully Connected Layers," in *2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, Hong Kong, China: IEEE, Jul. 2019, pp. 103–107. doi: 10.1109/AIM.2019.8868602.

[47] Y. Huang, Z. Tang, and K. Su, "Balance-batch: An Optimized Method for Semantic Segmentation Loss Functions," in *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Chongqing, China: IEEE, Jul. 2020, pp. 342–346. doi: 10.1109/CVIDL51233.2020.00-74.

[48] A. A. Lydia and F. S. Francis, "Convolutional Neural Network with an Optimized Backpropagation Technique," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India: IEEE, Mar. 2019, pp. 1–5. doi: 10.1109/ICSCAN.2019.8878719.

[49] Z. Farhadi, H. Bevrani, and M.-R. Feizi-Derakhshi, "Combining Regularization and Dropout Techniques for Deep Convolutional Neural Network," in *2022 Global Energy Conference (GEC)*, Batman, Turkey: IEEE, Oct. 2022, pp. 335–339. doi: 10.1109/GEC55014.2022.9986657.

[50] N. Bosso, M. Magelli, R. Trinchero, and N. Zampieri, "Application of machine learning techniques to build digital twins for long train dynamics simulations," *Veh. Syst. Dyn.*, pp. 1–20, Feb. 2023, doi: 10.1080/00423114.2023.2174885.

[51] M. D. Nirmal, P. P. Jadhav, and S. Pawar, "Pomegranate Leaf Disease Detection Using Supervised and Unsupervised Algorithm Techniques," *Cybern. Syst.*, pp. 1–12, Jan. 2023, doi: 10.1080/01969722.2023.2166192.

[52] F. Zhang, D. Li, S. Li, W. Guan, and M. Liu, "A Lightweight Tire Tread Image Classification Network," in *2022 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, Suzhou, China: IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/VCIP56404.2022.10008894.

[53] N. Martin, Z. Prince, V. Labouré, and M. Tano-Retamales, "Deep Learning for Multigroup Cross-Section Representation in Two-Step Core Calculations," *Nucl. Sci. Eng.*, vol. 197, no. 7, pp. 1406–1435, Jul. 2023, doi: 10.1080/00295639.2022.2159220.

[54] L. Zhuang, L. Gao, L. Ni, and B. Zhang, "An improved Expectation Maximization algorithm for hyperspectral image classification," in *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Gainesville, FL, USA: IEEE, Jun. 2013, pp. 1–

4. doi: 10.1109/WHISPERS.2013.8080631.

[55] J. Zhang, C. Yan, and X. Gong, “Subject dependent feature extraction method for motor imagery based BCI using multivariate empirical mode decomposition,” in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, Xiamen: IEEE, Oct. 2017, pp. 1–5. doi: 10.1109/ICSPCC.2017.8242378.

[56] H. Benbrahim and A. Behloul, “Fine-tuned Xception for Image Classification on Tiny ImageNet,” in *2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP)*, El Oued, Algeria: IEEE, Nov. 2021, pp. 1–4. doi: 10.1109/AI-CSP52968.2021.9671150.

[57] A. Alem and S. Kumar, “Deep Learning Models Performance Evaluations for Remote Sensed Image Classification,” *IEEE Access*, vol. 10, pp. 111784–111793, 2022, doi: 10.1109/ACCESS.2022.3215264.

[58] K. Murata, M. Mito, D. Eguchi, Y. Mori, and M. Toyonaga, “A Single Filter CNN Performance for Basic Shape Classification,” in *2018 9th International Conference on Awareness Science and Technology (iCAST)*, Fukuoka: IEEE, Sep. 2018, pp. 139–143. doi: 10.1109/ICAwST.2018.8517219.

[59] Q. Zhou, X. Zhang, and Y.-D. Zhang, “Ensemble Learning With Attention-Based Multiple Instance Pooling for Classification of SPT,” *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 69, no. 3, pp. 1927–1931, Mar. 2022, doi: 10.1109/TCSII.2021.3124165.

[60] M. Hersche, G. Karunaratne, G. Cherubini, L. Benini, A. Sebastian, and A. Rahimi, “Constrained Few-shot Class-incremental Learning,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 9047–9057. doi: 10.1109/CVPR52688.2022.00885.

[61] W. Liu, Y. Rao, B. Fan, J. Song, and Q. Wang, “Flower Classification using Fusion Descriptor and SVM”.

[62] J. D.Pujari, R. Yakkundimath, and Abdulmunaf. S. Byadgi, “SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 3, no. 7, p. 6, 2016, doi: 10.9781/ijimai.2016.371.