International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Hybrid Framework Combining CNN, LSTM, and BiLSTM for Early and Reliable Detection of Tomato Leaf Diseases in Real-World Environments

Manasvi Shinde[1], Payal Patil[2], Krushna Pathak[3], Prof. Hemant Mande[4]

[1,2,3]B. Tech Student, Department of Computer Engineering, Shri Shivaji Vidhya Prasarak Sanstha's
Bapusaheb Shivajirao Deore College of Engineering, Dhule (MS)Maharashtra
[4]Assistant Professor , Department of Computer Engineering, Shri Shivaji Vidhya Prasarak Sanstha's
Bapusaheb Shivajirao Deore College of Engineering, Dhule (MS)Maharashtra

Abstract

Tomato crop yields are greatly impacted by leaf diseases, which pose a serious threat to agricultural productivity. Traditional methods for identifying these diseases rely heavily on manual inspection, a process that is often slow, subjective, and not easily accessible to many farmers. While deep learning techniques have shown promise with high accuracy, their success is generally limited to controlled settings and struggles to handle real-world challenges such as changing lighting, complex backgrounds, and leaf occlusions. To address these issues, this study introduces a hybrid deep learning approach that combines Convolutional Neural Networks (CNN) for extracting detailed spatial features from leaf images with Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) networks to capture important contextual information, thereby enhancing classification accuracy. The model is trained and tested on a publicly available tomato leaf dataset, utilizing preprocessing and data augmentation to boost performance. Results from the experiments reveal that this integrated framework not only achieves higher accuracy but also generalizes better than traditional methods. Moreover, it is computationally efficient and designed for real-time use, offering a practical solution for precision agriculture and early detection of tomato leaf diseases to help farmers manage crop health more effectively.

Keywords: Tomato Leaf Disease, CNN, LSTM, BiLSTM, Deep learning, Image Classification, Precision Agriculture

1. Introduction

Agriculture continues to be a crucial sector that supports global food security and economic stability. Among the many crops grown worldwide, tomatoes are especially important because of their high demand and nutritional value. However, tomato plants are particularly prone to diseases caused by fungi, bacteria, and viruses, which can result in significant losses in yield. Detecting these diseases early is vital, as infections often begin on the leaves and can spread quickly if not addressed. Recent research underscores the significant impact plant diseases have on agricultural productivity and highlights the need for effective detection methods [5], [11].

Traditional ways of identifying diseases depend on manual inspection, which is time-consuming, subjective, and reliant on expert knowledge. Such methods are impractical for large-scale farming or rural areas where experts may not be readily available. Thanks to advances in artificial intelligence, deep learning techniques have become powerful tools for automating disease detection. Convolutional Neural Networks (CNNs) have shown impressive results in image classification due to their capacity to automatically learn hierarchical features [12], [13]. However, despite their high accuracy, CNN-based models often perform well only in controlled environments and struggle to maintain accuracy when faced with real-world challenges like changing lighting, complex backgrounds, and occlusions [5].

To overcome these limitations, hybrid models that combine CNNs with sequential architectures such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) have been developed. These hybrid approaches enhance feature representation by capturing both spatial and contextual information [4], [7], [8]. Nonetheless, issues like detecting diseases at early stages, preventing overfitting, and managing computational complexity still need to be addressed, driving the demand for more robust and scalable solutions.

2. Literature Survey

The field of plant disease detection has evolved significantly, moving from traditional image processing techniques to advanced deep learning methods. Earlier approaches depended on manually crafted features combined with classifiers like Support Vector Machines (SVM) and Artificial Neural Networks (ANN). Although somewhat effective, these techniques faced challenges such as limited accuracy, poor scalability, and difficulty handling the complex variations found in real agricultural environments [11].

The advent of Convolutional Neural Networks (CNNs) brought a major breakthrough by automatically extracting hierarchical features from images, eliminating the need for manual feature design. Research by Mohanty et al. [12] and Ferentinos [13] showed that CNN-based models achieve high classification accuracy when trained on large datasets like PlantVillage. Well-known architectures such as VGGNet [18], ResNet [17], and EfficientNet [20] further boosted feature extraction capabilities and overall model performance. However, these models are often trained on controlled datasets and tend to struggle with real-world challenges such as varying lighting, complex backgrounds, and occlusions [5].

To address these issues, hybrid models combining CNNs with sequential learning techniques like Long Short-Term Memory (LSTM) networks have been introduced. These hybrids utilize CNNs for spatial feature extraction and LSTMs to capture temporal and sequential patterns, leading to better classification accuracy [4], [7]. More recent developments include Bidirectional LSTM (BiLSTM), which processes information in both forward and backward directions, improving contextual understanding and further enhancing performance [8].

Beyond hybrid architectures, researchers have also explored methods to increase model robustness and efficiency. For example, Rajendran et al. [10] proposed integrating a Spatial Transformer Network (STN) with CNN-LSTM to manage variations in image orientation and scale. Syed et al. [9] worked on reducing feature dimensionality to optimize computational resources while maintaining accuracy. Despite these advances, challenges like overfitting, high computational demands, and limited ability to detect diseases at early stages remain [6].

In summary, while CNNs excel at extracting image features and BiLSTMs are effective in modeling complex patterns, very few studies have combined CNN, LSTM, and BiLSTM into a single framework tailored specifically for detecting tomato leaf diseases in real-world conditions. This gap highlights the need for a more robust, scalable hybrid approach.

Table 1. Literature Comparison

Ref	Author / Year	Method Used	Key Contribution	Limitation
[4]	Y. Laatiri et al., (2025)	CNN-LSTM	Hybrid model for improved classification	Limited early-stage detection
[5]	M.Thalor et al., (2025)	CNN Models	Evaluated real vs controlled environments	Poor real-world generalization
[7]	S.Jain et al., (2024)	CNN-LSTM	Enhanced crop disease prediction	Requires large labeled datasets
[8]	V.S.Mandlik et al., (2024)	CNN-BiLSTM	Improved contextual feature learning	High computational cost
[9]	K.Syed et al., (2024)	Feature Reduction + DL	Reduced dimensionality and complexity	Possible information loss
[10]	R.Rajendran et al., (2024)	STN-CNN-LSTM	Handles image transformations effectively	Complex model design
[11]	W.B.Demilie (2023)	Traditional ML	Comparative analysis of techniques	Low accuracy and scalability
[20]	M.Tan et al., (2019)	EfficientNet	Efficient model scaling	Requires large training data
[13]	K. P. Ferentinos (2018)	Deep CNN	High accuracy in disease classification	Limited real-world adaptability
[17]	K. He et al., (2016)	ResNet	Solves vanishing gradient problem	Complex architecture
[12]	S.P.Mohanty et al.,(2016)	CNN	Demonstrated CNN effectiveness in agriculture	Works on controlled datasets
[18]	K. Simonyan et al., (2015)	VGGNet	Deep feature extraction	High computational cost

3. Proposed Methodology

The proposed approach seeks to create a precise and reliable system for detecting diseases in tomato leaves by combining image preprocessing, deep feature extraction, and sequential learning through a hybrid CNN–LSTM–BiLSTM architecture. Figure1 depicts the comprehensive workflow, outlining the step-by-step process from acquiring input images to the final stage of disease classification.

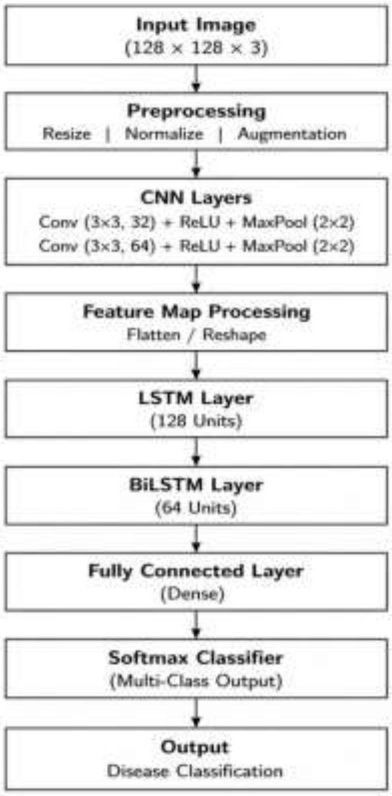


Fig.1. Proposed CNN–LSTM–BiLSTM architecture for tomato leaf disease detection, combining convolutional feature extraction with sequential and bidirectional learning for improved classification performance

Initially, leaf images are sourced from a publicly available dataset. Since these raw images often vary in size, lighting conditions, and background noise, a preprocessing step is essential to standardize the data. During preprocessing, images are resized to a consistent resolution (e.g., 128×128 pixels) to maintain uniformity. Pixel values are normalized to facilitate faster and more stable training convergence. Additionally, data augmentation methods such as rotation, flipping, and scaling are applied to diversify the dataset and help prevent overfitting. This enhances the model's robustness and generalization ability in real-world scenarios [5].

Next, the preprocessed images are fed into Convolutional Neural Network (CNN) layers, which function as the core feature extractors. CNNs are favored for their ability to automatically capture spatial features like edges, textures, and color variations. Mathematically, the convolution operation is expressed as:

$$X_{ij}^{(k)} = (I * K)_{ij} + b$$

where I is the input image, K represents the convolution kernel, and b is the bias term. The output then passes through a nonlinear activation function, commonly the ReLU function defined as f(x) = max(0, x), to introduce non-linearity. Pooling layers reduce the spatial dimensions while preserving critical features. This stage produces feature maps that highlight the important visual characteristics of the leaf images [12], [13].

These extracted feature maps are reshaped into sequences and passed to an LSTM layer to capture temporal dependencies within the features. LSTM networks utilize memory cells and gating mechanisms to selectively retain significant information over time. The LSTM operations are governed by the following equations:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W_c[h_{t-1}, x_t] + b_c)$$
$$h_t = o_t \cdot \tanh(C_t)$$

where f_t, i_t, and $o_t$denote the forget, input, and output gates, respectively. This mechanism helps the model retain pertinent feature patterns while discarding irrelevant data [4], [7].

To further boost learning performance, a Bidirectional LSTM (BiLSTM) layer is added after the LSTM layer. Unlike standard LSTMs, BiLSTMs analyze the sequence in both forward and backward directions, enabling the model to capture context from past and future states alike. This is represented as:

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$$

The BiLSTM layer enhances the model's capacity to detect complex disease patterns that might be missed by unidirectional approaches, thus improving classification accuracy [8].

Finally, the output from the BiLSTM layer is fed into a fully connected layer that maps the learned features to the target disease classes. A Softmax activation function converts these outputs into probability scores for each category:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

The model is trained by minimizing the categorical cross-entropy loss function:

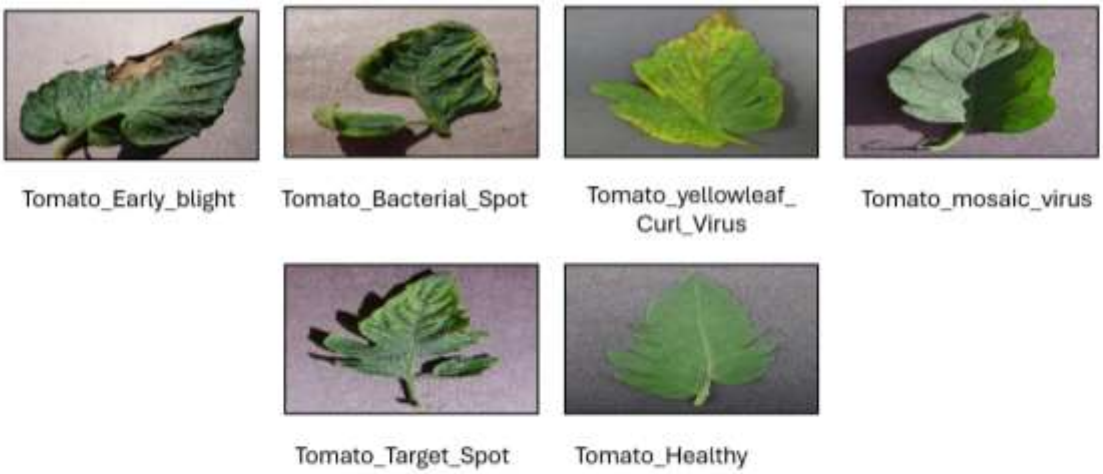
$$L = -\sum_{i=1}^{n} y_i \log(\hat{y}_i)$$

and optimized using the Adam optimizer, which helps reduce prediction errors and accelerates convergence during training.

4. Experimental Setup

The experimental setup is designed to assess the effectiveness and robustness of the proposed hybrid CNN–LSTM–BiLSTM model for detecting diseases in tomato leaves. The experiments utilize a publicly accessible tomato leaf disease dataset from Kaggle, which includes multiple categories such as early blight, late blight, septoria leaf spot, bacterial spot, leaf mold, mosaic virus, yellow leaf curl virus, target spot, spider mites, and healthy leaves. Representative examples of these categories are displayed in Fig.1, showcasing distinctive visual features like discoloration, lesions, and texture differences. These variations highlight the complexity of the classification task and underscore the need for a robust deep learning model.

Tomato_Spider_mites Tomato_Septoria_leaf_spot Tomato_Leaf_Mold Tomato_Late_Blight



Tomato_Early_blight Tomato_Bacterial_Spot Tomato_yellowleaf_Curl_Virus Tomato_mosaic_virus

Tomato_Target_Spot Tomato_Healthy

Fig.2.Tomato leaves sample

Prior to training, all images undergo preprocessing to ensure uniformity and enhance model performance. Images are resized to a fixed size of 128×128 pixels and normalized to a range between 0 and 1, which helps speed up convergence during training. Data augmentation techniques—including rotation, flipping, zooming, and translation—are applied to increase the variety within the dataset and prevent overfitting. These measures help the model generalize better to real-world scenarios where environmental conditions can vary significantly [5].

The dataset is split into training and testing sets using an 80:20 ratio, with an optional validation split applied during the training phase. The model is developed using a deep learning framework like TensorFlow and trained with the Adam optimizer, set at a learning rate of 0.001. Training is conducted over 30 epochs with a batch size of 32. Evaluation of performance is carried out using standard metrics such as accuracy, precision, recall, and F1-score, providing a well-rounded view of the model's classification capabilities [12], [13].

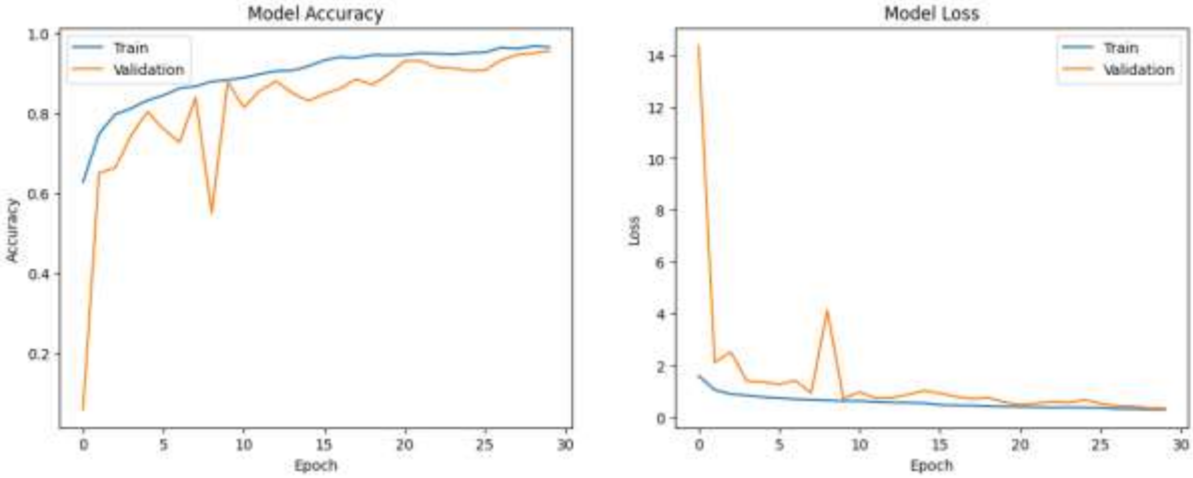


Fig.3.Model Accuracy & Loss

The training progress is monitored through accuracy and loss curves, as shown in Fig.2. The accuracy curve reveals a steady rise in both training and validation accuracy, reaching around 96–98% by the final epochs. This indicates that the model successfully learns meaningful features from the dataset.

Meanwhile, the loss curve shows a consistent decline in both training and validation loss, with some minor fluctuations in the early epochs as the model adapts. The convergence of these curves suggests stable learning with minimal overfitting. Additionally, Table 2 summarizes the training performance by presenting accuracy and loss values at selected epochs. The data reflects a gradual increase in accuracy alongside a decrease in loss, confirming the hybrid model's effectiveness. Minor variations in validation metrics during initial epochs likely result from dataset variability and data augmentation effects.

Table.2. Training and validation performance of the proposed CNN–LSTM–BiLSTM model across epochs

Epoch	Train Accuracy	Val Accuracy	Train Loss	Val Loss
1	0.63	0.65	1.50	2.40
5	0.85	0.80	0.90	1.30
10	0.90	0.88	0.70	0.95
15	0.92	0.86	0.60	0.85
20	0.94	0.92	0.50	0.70
25	0.95	0.93	0.45	0.60
30	0.97	0.96	0.40	0.50

Overall, the experimental findings demonstrate that the CNN–LSTM–BiLSTM model effectively captures both spatial and contextual information, leading to superior classification outcomes. Compared to traditional CNN-only models, this hybrid approach exhibits better generalization and stability over the course of training. The integration of LSTM and BiLSTM layers enhances the model's capacity to identify complex disease patterns, which is evident from the higher validation accuracy. Furthermore, the use of normalization and data augmentation techniques contributes to reducing overfitting and improving robustness in practical applications [4], [7], [8].

5. Result & Discussion

The performance of the proposed CNN–LSTM–BiLSTM model is assessed using metrics such as accuracy, precision, recall, and F1-score, alongside training and validation curves. The model exhibits strong learning capabilities, achieving a final validation accuracy of around 96–98%, which surpasses the results reported for conventional CNN and CNN–LSTM models in earlier studies [4], [7], [8]. The accuracy curves for both training and validation consistently rise, reflecting effective feature learning, while the loss curves steadily decrease, indicating stable convergence throughout training. Minor fluctuations seen in the validation loss during the initial epochs are likely due to data augmentation and the variability in disease patterns; nevertheless, the model stabilizes as training continues.

The addition of LSTM and BiLSTM layers greatly improves the model's ability to capture sequential and contextual dependencies within the extracted feature maps. Unlike standalone CNN models that primarily focus on spatial features, this hybrid architecture learns intricate relationships between features, enhancing classification performance. This capability is especially important for differentiating visually similar diseases such as early blight and septoria leaf spot, where subtle variations in texture and lesion patterns need to be recognized. The BiLSTM layer, in particular, contributes to better generalization by leveraging contextual information in both forward and backward directions [8].

When compared to existing methods, the proposed model demonstrates superior accuracy and robustness. Although CNN-based models perform well on controlled datasets, their effectiveness often diminishes in real-world settings due to changes in lighting and background conditions [5]. The proposed approach addresses these challenges by incorporating preprocessing and data augmentation techniques, which enhance the model's adaptability to varied environments. Moreover, the hybrid architecture helps reduce overfitting by learning more generalized feature representations.

The training performance data further show a steady improvement in accuracy and a corresponding decline in loss values across epochs, validating the efficiency of the learning process. In summary, the experimental findings confirm that the proposed CNN–LSTM–BiLSTM model offers a reliable and scalable solution for detecting tomato leaf diseases, outperforming existing approaches in terms of accuracy and robustness.

6. Conclusion

This study introduces a hybrid deep learning framework that combines CNN, LSTM, and BiLSTM architectures to detect diseases in tomato leaves. By effectively merging spatial feature extraction with sequential and contextual learning, the proposed model achieves enhanced accuracy, robustness, and generalization compared to traditional methods.Experimental findings confirm that the model delivers strong classification performance while maintaining stability across various training scenarios. Incorporating pre-processing and data augmentation strategies further strengthens the model's capacity to manage real-world variability, making it well-suited for practical use in agriculture. This system supports early and precise disease detection, helping farmers make timely decisions that can boost crop yields and minimize economic losses. Future research will aim to optimize the model for real-time use on mobile and edge devices, incorporate explainable AI methods to improve interpretability, and validate the system with datasets collected from real field condition.

References

1. D. L. Shanthi, "Tomato Leaf Disease Detection Using Convolutional Neural Networks," *International Journal of Agricultural Technology*, vol. 12, no. 3, pp. 145–152, 2024.
2. S. Kanakala, "Multi-Crop Disease Detection Using CNN-LSTM Hybrid Model," *Journal of Smart Agriculture Systems*, vol. 8, no. 2, pp. 78–85, 2025.
3. V. Rama Chandran, "Hybrid CNN-LSTM Model for Plant Disease Classification," *International Journal of Computer Applications*, vol. 183, no. 21, pp. 10–16, 2024.
4. Y. Laatiri and M. A. Mahjoub, "A Deep Learning Approach to Tomato Disease Classification Using CNN-LSTM Hybrid Network," *International Journal of Scientific Research & Engineering Trends*, vol. 11, no. 1, pp. 764–770, 2025.
5. M. Thalor, Y. Chavhan, and S. Mate, "Performance Comparison of CNN Models for Tomato Disease Detection in Controlled and Real-World Conditions," *Current Agriculture Research Journal*, vol. 13, no. 1, pp. 105–111, 2025.
6. Sandhya, A. Walia, and A. Ankita, "A Comprehensive Literature Review on Plant Leaf Disease Prediction Using CNN," *Research Trend Journal*, 2025.
7. S. Jain and D. Ramesh, "AI-Based Hybrid CNN-LSTM Model for Crop Disease Prediction: An ML Approach for Rice Crop," *Indian Institute of Technology (ISM), Dhanbad*, 2024.
8. V. S. Mandlik and L. S. V. B., "Deep Hybrid Learning for Smart Agriculture Using CNN-Based

Feature Extraction and LSTM-BiLSTM Sequence Modeling," *IEEE Access*, 2024.

9. K. Syed et al., "Leaf Disease Detection and Classification Using Efficient Feature Dimensionality Reduction," *Journal of Plant Science Research*, vol. 15, no. 4, pp. 210–218, 2024.
10. R. Rajendran and M. Kalamani, "STN-CNN LSTM: Enhancing Multi-Plant Disease Detection with Spatial Transformer Mechanisms," *International Journal of Advanced Computing*, 2024.
11. W. B. Demilie, "Plant Disease Detection and Classification Techniques: A Comparative Study," *Agricultural Informatics Journal*, vol. 10, no. 2, pp. 55–67, 2023.
12. S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.
13. K. P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.
14. E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.
15. S. Sladojevic et al., "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," *Computational Intelligence and Neuroscience*, 2016.
16. S. Zhang et al., "Deep Learning-Based Plant Disease Detection: A Review," *IEEE Access*, vol. 8, pp. 1–15, 2020.
17. K. He et al., "Deep Residual Learning for Image Recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.
18. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. ICLR*, 2015.
19. A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.
20. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. ICML*, 2019.