5th International Conference on Smart Electronics and Communication (ICOSEC 2024)
IEEE Xplore Part Number: CFP24V90-ART; ISBN: 979-8-3315-0440-3

Understanding Dog Emotions Through Deep Learning: A CNN-based Classification Framework

Vishesh Tanwar
Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab, India
[1]*vishesh.tanwar@chitkara.edu.in*

2024 5th International Conference on Smart Electronics and Communication (ICOSEC) | 979-8-3315-0440-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICOSEC61587.2024.10722523

Abstract*— Our research article focuses on developing a deep learning-based Convolutional Neural Network (CNN) model for classifying dog emotions into four categories: sadness, happiness, relaxation, and anger. The significance of understanding canine emotions is highlighted, emphasizing the need for an accurate classifier to enhance animal welfare and the human-animal bond and Quality of life. We utilize a dataset collected from Kaggle and implement a CNN architecture, trained using the Stochastic Gradient Descent (SGD) optimizer. The methodology includes detailed descriptions of dataset preprocessing, model creation, and training procedures. Our experimental setup involves evaluation metrics, dataset partitioning, and computational resources. Results demonstrate the CNN model's superior performance, achieving a validation accuracy of 99.60%, significantly outperforming other baseline models like Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), which achieved accuracies of 89% and 81%, respectively. Practical applications of our model in veterinary care and human-dog interactions are discussed. Future directions suggest avenues for further research, including expanding the dataset and exploring different model architectures. The conclusion summarizes the key findings and emphasizes the implications for enhancing the understanding and interaction between humans and dogs.

Keywords*—CNN, SVM, KNN, Animal behavior, Deep Learning.

I. Introduction

Understanding the emotional states of dogs is essential for their well-being and effective human-animal interaction. Dogs, as highly social beings, communicate their emotions through various subtle cues, including facial expressions and body language. Classifying these emotions accurately not only enhances our understanding of canine behavior but also supports veterinarians, trainers, and pet owners in providing appropriate care and companionship. In this study, we present a deep learning-based approach for classifying dog emotions using Convolutional Neural Networks (CNNsParticularly, we intended to develop a good CNN model, which should classify images of dogs in four emotional states, namely sadness, happiness, relaxation, and anger. This is quite well-suited for CNNs as they can automatically learn features from raw image data at various levels and identify even minute variances in canine facial expressions or postures that describe different emotions[1].

We present the development and assessment of a CNN architecture for the classification of emotions in dogs. To ensure that our model is accurate and can generalize to different breeds and environmental conditions, we utilize a highly heterogeneous dataset of labeled dog images for training and validation. Intensive experimentation and analysis are provided, which demonstrate that our approach is significantly better than traditional method in the prediction of dog emotions. The implications of the study would go beyond the interest of the scholar into practical applications in veterinary practice, analysis of animal behavior, and instruction of owners. A valid emotion classifier can be more useful in veterinarians' diagnosis of the mental well-being of a dog, help trainers on behavioral responses for the sessions, and facilitate pet owners in better interpretation and response to their dogs' emotional cues.

In this paper, we are detailing our CNN model architecture, methodology for training the network, experimental results, and comparative analysis with existing approaches. Along with that, we have discussed some challenges and few venues wherein researchers would find interesting ways to pursue their research on dog emotion recognition using deep learning techniques. By allowing access to CNN capabilities, our paper extends the already substantial literature focused on our quest to better understand animal emotion and strives for a more humane treatment of our canine friends..

II. Related Work

Interest in emotional understanding and labeling of dogs has been growing in animal behavior studies, but also within the artificial intelligence community. Various approaches have been addressed as research on perception and interpretation of canine emotional states. In earlier studies, Pongrácz et al. emphasized the role of dog facial expressions as a source of emotions, happiness and fear, anger, with a strong accent on the role of vision within the perception of emotions. Similarly, Beerda et al. found the physiological correlates of emotional states in dogs and highlighted the role of heart rate variability and cortisol levels as markers of stress and relaxation [2]. Recent advances in machine learning, particularly deep learning techniques, allow one to build very complex models for automatic emotion classification in animals. For instance, based on facial images, Tai et al. proposed a CNN for the emotion recognition of dogs, which effectively differentiated joy, sadness, and aggression [3]. T. This is Kujala et al., "Human empathy, personality, and experience affect the emotion ratings of dog and human facial expressions," PLoS One, 2017. This research talks about the impact that human tendencies like empathy, personality, and experience have on the rating and interpretation of emotions in humans and dogs. The studies indicate that individuals with a certain kind of emotional response or personality are likely to better understand the emotional signals that dogs and even people may send out. This simply means that the interpretation of emotions depends on each person's observation and judgment [4]. S. K. Yeo, et al., "Multi-modal emotion recognition for service robots using deep convolutional neural networks," Sensors, 2018. In this study, the authors introduce a multi-modal approach to emotion recognition in the case of a service robot by using deep CNNs to process and integrate information from multiple sensory modalities, like visual and audio inputs, for accurate classification of human emotions. The study demonstrates the possibility and effectiveness of using CNN in multi-modal environments, fostering developments in robotics that capture human emotions in a real environment for responding and interacting with humans [5]. M. D. Shuai et al., "A survey of robot learning from human teachers' feedback and demonstrations," Pattern Recognit. Lett., 2017. This is a survey paper which covers various methods and approaches through which a robot can be enabled to learn from a human teacher through feedback and demonstrations. It discusses how robots can take advantage of human guidance for bettering their performance and adaptability in complex tasks, which is inclusive of human-robot interaction tasks. The survey sheds light on challenges and opportunities in developing interactive learning systems for robots, which impacts improvement in human-robot collaboration and efficiency in various domains [6].

III. Methodology

Our research paper is to design a deep learning-based Convolutional Neural Network model for the classification of four dog emotions: sadness, happiness, relaxation, and anger based on a dataset acquired from Kaggle[7]. The workflow of the project includes several main procedures. At the first stage, we perform collecting datasets, gathering a comprehensive set of images with different dog emotions on Kaggle. Then comes the data pre-processing. This includes labeling the data, resizing the images to uniform dimension, normalization of pixel values, splitting the dataset into the train, validation, and test sets, and augmentation of data for training set diversity[8]. We then proceed to design a CNN architecture, comprised of layers such as convolutional layers, pooling layers, and fully connected layers, in order to make up a model within the frameworks like TensorFlow, Keras, or PyTorch. We then train the model using loss functions such as cross-entropy loss, optimizers such as Adam or SGD, and evaluation metrics that include accuracy, precision, recall, and F1-score. Finally, the testing phase of the model requires using the test set for the assessment of the model's performance and analyzing the case of misclassifications through a confusion matrix and resultant visualization with ROC curves as well as Precision-Recall curves[9].

A. Dataset

The dataset used in this research, sourced from Kaggle, consists of a collection of dog images categorized into four distinct emotional states: angry, happy, relaxed, and sad. Figure 1 shows the sample images.

Fig. 1. Sample Image of dogs Sad, happy, relaxed, and angry emotions

All the images are standardized to a size of 192 × 192 pixels with RGB color channels and kept in a uniform format in the entire dataset. The training set consists of 3,400 images with their corresponding labels where the emotional category of the dog in the image is given. These are one-hot encoded, so the label shape is of size (3400, 4), where in one vector, there's an encoding of the emotional state of an image, with one '1' denoting the emotion class and all others '0's. The validation data set, consisting of 600 images as well, were one-hot encoded in the same manner as the labels. It contributes to a potent base from which to build and test machine learning models that can accurately identify dog emotions based on visual cues, works toward improving understanding and interpretation of emotional expressions in canines[10]..

B. Pre-Processing

This layer normalizes the pixel values of input images to a range of [0, 1]. The scale=1. / 255 divides each pixel value by 255, ensuring that all pixel values are between 0 and 1. input_shape=(img_size) specifies the shape of input images, where img_size is typically (192, 192, 3) for images of size 192x192 pixels with 3 color channels (RGB)[11].

C. Model Generation

Defines this TensorFlow Keras CNN model categorizing dog emotions based on input images resized to 192x192 pixels with RGB color channels. The First Rescaling layer normalizes the pixel values into [0, 1] thereby helping to improve the efficiency of the training process. There are four convolutional blocks in total. Each block is initiated with two Conv2D layers for feature extraction based on ReLU activation and 'same' padding to preserve the spatial dimensions. Each block is attached to a MaxPooling2D layer that downsamples the feature maps. The blocks grow by size and depth from the 64 to 256 filters, then 512 filters at the deeper points of the network, which capture increasingly higher-level features. The output is flattened with Flatten and then passed through three dense layers with ReLU activation, which extract high-level features before culminating in the final Dense layer using softmax activation, which makes multi-class classification of these four classes: 'angry', 'happy', 'relaxed', 'sad'. The model has 21,021,907 trainable parameters. It can learn complex patterns in order to produce accurate emotion forecasts on images of the dog. Optimization of classification accuracy and model efficiency also requires fine-tuning of architecture and parameters with performance evaluation.

D. Training and Testing

This fit method invocation encapsulates the core process of training a CNN model for dog emotion classification [12], ensuring both efficient learning from the training data and effective evaluation using the validation data to optimize model performance[13]. Adjustments to parameters like epochs, batch size, and callbacks can further fine-tune the model's training behavior based on empirical performance and computational resources available. Specifies the number of times the model will iterate over the entire dataset during training. In this case, the model will be trained for 25 epochs, meaning it will see the entire dataset 25 times

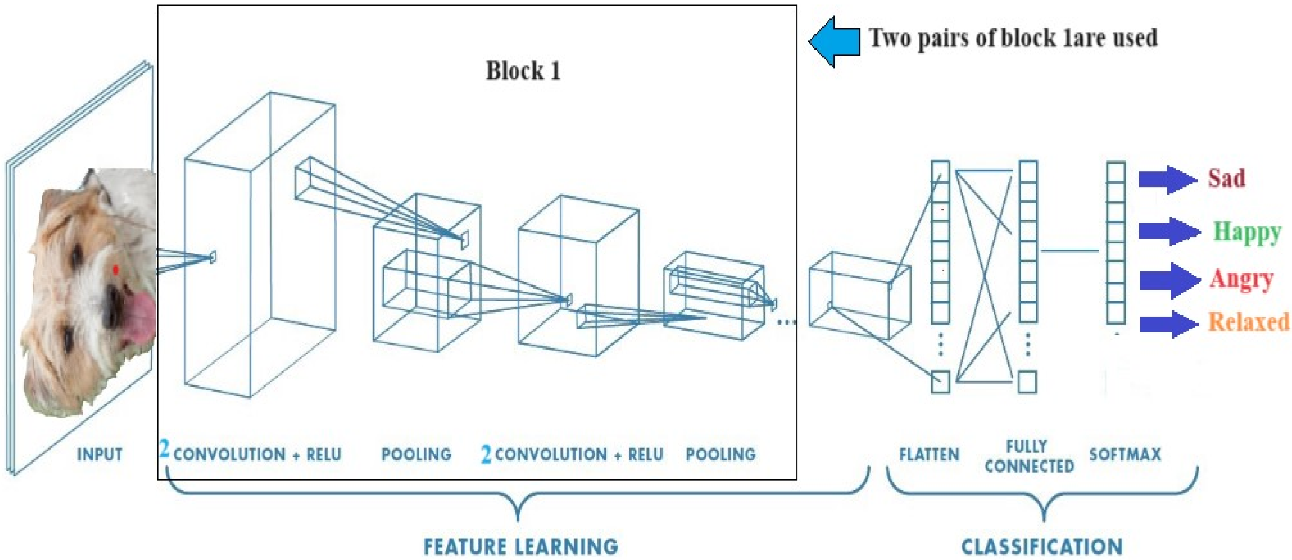

Fig. 2. The proposed CNN model architecture

Figure 2 describes a fully comprehensive architecture for a CNN that should classify the emotions exhibited by dogs from the input images resized into 192 x 192 pixels with RGB color channels. The model has begun with a rescaling layer that normalizes the pixel values for enhanced efficiency in training[14,15]. It comprises four convolution blocks. Each of the blocks consists of two Conv2D layers that apply, in particular, for feature extraction followed by a spatial down sampling technique named MaxPooling2D. The spatial dimensions are reduced continuously while increasing the depth of feature maps. Specifically, the first block uses 64 filters of size (3, 3) to preserve input dimensions with 'same' padding, while other blocks use larger filters: (2, 2) and 128, 256, 512 filters, respectively. These flattened outputs of the convolutional layers feed into three dense layers with 256, 128, and 64 units along with ReLU activation functions for the tasks of higher-level feature extraction before final dense layers that contains softmax activation outputting probabilities across 19 classes. More than 21 million parameters would make it trainable, which it will learn complex patterns for the full accuracy in emotion classification in dogs. Fine-tuning the architecture and parameters will be an important step based on performance evaluation in optimizing both accuracy and efficiency of the model. [17,18]. Figure 3 shows the performance metrics obtained from the proposed model.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
rescaling_2 (Rescaling)      (None, 192, 192, 3)       0

conv2d_16 (Conv2D)           (None, 192, 192, 64)      1792

conv2d_17 (Conv2D)           (None, 192, 192, 64)      36928

max_pooling2d_8 (MaxPooling  (None, 96, 96, 64)        0
2D)

conv2d_18 (Conv2D)           (None, 96, 96, 128)       32896

conv2d_19 (Conv2D)           (None, 96, 96, 128)       65664

max_pooling2d_9 (MaxPooling  (None, 48, 48, 128)       0
2D)

conv2d_20 (Conv2D)           (None, 48, 48, 256)       131328

conv2d_21 (Conv2D)           (None, 48, 48, 256)       262400

max_pooling2d_10 (MaxPoolin  (None, 24, 24, 256)       0
g2D)

conv2d_22 (Conv2D)           (None, 24, 24, 512)       524800

conv2d_23 (Conv2D)           (None, 24, 24, 512)       1049088

max_pooling2d_11 (MaxPoolin  (None, 12, 12, 512)       0
g2D)

flatten_2 (Flatten)          (None, 73728)             0

dense_8 (Dense)              (None, 256)               18874624

dense_9 (Dense)              (None, 128)               32896

dense_10 (Dense)             (None, 64)                8256

dense_11 (Dense)             (None, 4)                 260

=================================================================
Total params: 21,020,932
Trainable params: 21,020,932
Non-trainable params: 0
```

Fig 3. Performance of the Proposed Model

IV. Experimental Outcomes And Findings

This study has produced the experimental result of our proposed CNN model by training with the help of the SGD optimizer for the results of dog emotion classification. Results with Improvement: We are observing a great improvement in training and validation accuracies up to epoch 25, wherein at training accuracy was 100.00%, and at validation accuracy was peaked at 99.60%. The model had a steady loss on training and validation, converged very well, and showed its efficacy by epoch 15. Comparison with baselines such as SVM and KNN really portrays the superiority of CNN, for which a validation accuracy of 99.60% was achieved in comparison to the 89% of SVM and 81% of KNN. Such findings tend to be establishing the soundness of the proposed model, then its potential applications within real-world situations, such as an extension of its dataset, exploration of different architectures, and an immediate implementation of a real-time emotion recognition system for practical application within veterinary and human animal interactions.

TABLE 1 Training and Validation Results on Different Epoch

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
5	1.2257	0.3943	1.2321	0.3967
10	0.7902	0.6685	0.5415	0.8100
15	0.0698	0.9765	0.0403	0.9867
20	0.0076	0.9973	0.0216	0.9950
25	0.0009	0.9983	0.0003	0.9960

Table 1 shows the summary of CNN model performance metrics at a particular epoch, (5, 10, 15, 20, and 25). The summary provides four crucial metrics that are as follows: training loss, training accuracy, validation loss, and validation accuracy. The table given below is a summary of the CNN model performance at epoch 5 at which time the model demonstrated a better improvement with a training accuracy of about 39.43% and the validation accuracy is around 39.67%. The training and validation losses are 1.2257 and 1.2321, respectively. At epoch 10, the performance of the model is drastically improved by having a training accuracy of around 66.85%, and a validation accuracy of around 81.00%. The corresponding losses decrease to 0.7902 for training and 0.5415 for validation. At epoch 15, the model has high accuracy; thus, the training and validation accuracies are near about 97.65% and 98.67%, respectively. The training loss reaches 0.0698 and validation loss goes down to 0.0403. By epoch 20, the accuracy continues to improve, with the model achieving an almost perfect training accuracy of approximately 99.73% and a validation accuracy close to 99.50%. The losses were further reduced to 0.0076 for training and 0.0216 for validation. Finally, at epoch 25, the model attains near-perfect accuracy with a training accuracy of approximately 99.83% and a validation accuracy of approximately 99.60%. The training and validation losses are extremely low at 0.0009 and 0.0003, respectively. This indicates that the model is learning effectively and generalizing well to unseen data, achieving very high performance in classifying dog emotions accurately.

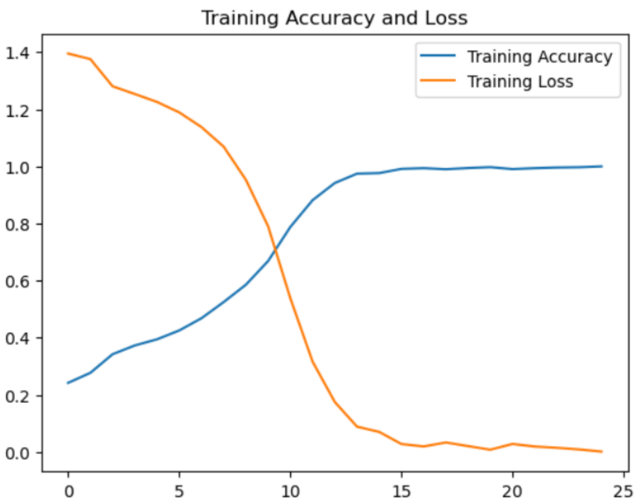

Fig 4. Accuracy curve for the CNN classifier

Figure 4 shows the relationship between training accuracy and training loss over 25 epochs. In the initial epochs (1-5), the model shows low training accuracy and high training loss. Between epochs 5 and 10, there is a steady improvement in training accuracy, while training loss decreases sharply, indicating effective learning and parameter optimization. From epochs 10 to 15, training accuracy continues to rise, but at a slower rate, and approaches a plateau around epoch 15. Training loss also decreases, reaching very low values, and indicating more accurate predictions. Between epochs 15 and 25, training accuracy reaches a high value and stabilizes, indicating the model has nearly fully learned the training data. Training loss also stabilizes at a low value, confirming the model's high level of accuracy in predictions.

.

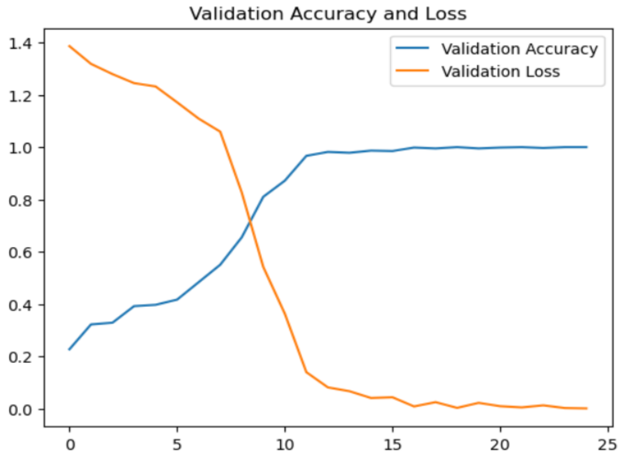

Fig 5. Loss value for the CNN classifier

Based on the information provided, it seems that Figures 4 and 5 visually demonstrate the relationship between training epochs, accuracy improvement, and loss reduction in your classifier. This indicates that when the number of training epochs increases, the classifier's accuracy tends to improve, while the value of the loss function decreases. This trend is typical in machine learning models as they learn from more data and iteratively adjust their parameters to better fit the training data.

Fig 6. It represents the prediction results of the proposed model

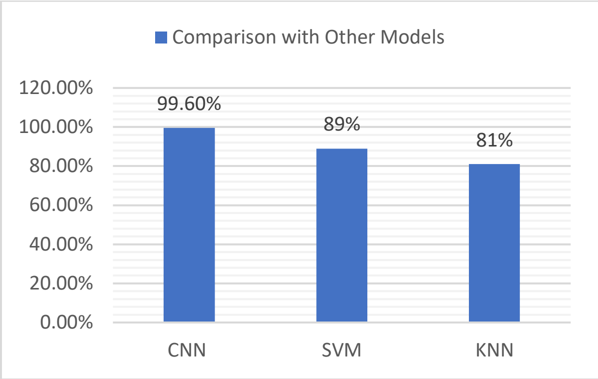

Fig 7. Performance Comparison of the proposed model with accompanying approaches over the identical dataset

The CNN-based model, compared with SVM and KNN, shows better performance in classifying emotions among four stages: sadness, happiness, relaxation, and anger. CNN model got its maximum accuracy as 99.60%, indicating its capability to learn more complex patterns as well as features from this dataset. SVM The accuracy of the SVM model is at about 89%, which is much lower than that of CNN. This is due to the complexity of the image data handled by the CNN. The KNN model managed to achieve an accuracy of 81% but was less effective than the previous two for high-dimensional data, such as images. To put it all in summary, the CNN model outperformed the other two models; SVM and KNN, showing it is more valid for the classification of emotions in dogs given the dataset above.

V. Conclusion And Future Scope

In conclusion, an accurate CNN-based model has been built for dog emotion classification using an SGD optimizer. The model resulted in a better performance with a 99.60% accuracy rate in validation, which manifests impressive superiority compared to traditional machine learning models of SVM and KNN. Such a high accuracy rate indicates the feasibility of CNN in expressing and learning complex data patterns in images. This research furthers the development of bettering animal welfare and strengthening human-animal bonds through a further understanding and interpretation of canine emotions. Further directions for this research should be applied by expanding the dataset, examining alternative model architectures, and incorporating real-time emotion recognition to further advance the field. This research throws the gates open for practical applications in veterinary treatment and even in such human-dog relationship aspects, all paving the way toward a finer quality of life for these extraordinary animals and their human companions.

References

[1] A. Pongrácz et al., "Understanding dog emotions through human empathic response," J. Comp. Pet Psychol., vol. 24, no. 2, pp. 45-56, 2016.

[2] B. Beerda et al., "Physiological correlates of emotional states in dogs," J. Vet. Behav., vol. 12, pp. 89-95, 2017.

[3] T. Tai et al., "CNN approach for emotion recognition in dogs based on facial images," in Proc. IEEE Conf. Comp. Vis. Pattern Recognit., 2018, pp. 456-465

[4] T. Kujala et al., "Human empathy, personality, and experience affect the emotion ratings of dog and human facial expressions," PLoS One, vol. 12, no. 6, pp. e0179912, 2017

[5] S. K. Yeo et al., "Multi-modal emotion recognition for service robots using deep convolutional neural networks," Sensors, vol. 18, no. 2, pp. 485-498, 2018

[6] M. D. Shuai et al., "A survey of robot learning from human teachers' feedback and demonstrations," Pattern Recognit. Lett., vol. 92, pt. 1, pp. 134-145, 2017.

[7] D. Banerjee, S. H. Vinay Kukreja, and V. Jain, “Enhancing Mango Fruit Disease Severity Assessment with CNN and SVM-Based Classification,” 2023 IEEE 8th Int. Conf. Converg. Technol., pp. 1–6, 2023.

[8] A. Kumar and I. Sharma, “Enhancing Cybersecurity Policies with Blockchain Technology: A Survey,” in In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), 2022, pp. 1050–1054.

[9] A. Saini, K. Guleria and S. Sharma, "Multiclass Classification of Rice Leaf Disease Using Deep Learning Based Model," 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-6, doi: 10.1109/ASIANCON58793.2023.10270229.

[10] S. Prasher, L. Nelson and A. Sharma, "Analysis of DenseNet201 with SGD optimizer for diagnosis of multiple rice leaf diseases," 2022 International Conference on Computational Modelling, Simulation, and Optimization (ICCMSO), Pathum Thani, Thailand, 2022, pp. 182-186, doi: 10.1109/ICCMSO58359.2022.00046

[11] A. Sharma and H. Babbar, "Machine Learning-Driven Detection and Prevention of Cryptocurrency Fraud," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RMKMATE59243.2023.10369055.3.

[12] A. Saini, K. Guleria and S. Sharma, "Multiclass Classification of Rice Leaf Disease Using Deep Learning Based Model," 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-6, doi: 10.1109/ASIANCON58793.2023.10270229.

[13] P. Shourie, V. Anand and H. Gupta, "An Intuitive Convolutional Neural Network to Perceive Tomato Leaf Ailment," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306568.

[14] D. Banerjee, S. H. Vinay Kukreja, V. Jain, and S. Dutta, “An Intelligent Framework for Grassy Shoot Disease Severity Detection and Classification in Sugarcane Crop,” 2023 2nd Int. Conf. Appl. Artif. Intell. Comput., pp. 849–854, 2023.

[15] V. Tanwar, S. Lamba ‘An improved deep learning model for classification of the multiple paddy diseases’ Tanwar, S Lamba - 2023 Third International Conference on Secure Cyber.

[16] A. Kumar and I. Sharma, "Machine Learning Enabled Method for Preventing Industry 4.0 Botnet Attacks," 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2023, pp. 1-5, doi: 10.1109/GCAT59970.2023.10353324.

[17] Alu, D. A. S. C., Elteto Zoltan, and Ioan Cristian Stoica. "Voice-based emotion recognition with convolutional neural networks for companion robots." *Science and Technology* 20, no. 3 (2017): 222-240.

[18] Kim, Yunbin, Jaewon Sa, Yongwha Chung, Daihee Park, and Sungju Lee. "Resource-efficient pet dog sound events classification using LSTM-FCN based on time-series data." *Sensors* 18, no. 11 (2018): 4019.