**AUTOMATED IDENTIFICATION OF INDIAN HERITAGE MONUMENTS USING VGG16-BASED CONVOLUTIONAL NEURAL NETWORKS**

**ABSTRACT**

Monument recognition is a challenging task in the domain of image classification. Different structure orientations play a significant part in recognizing monuments in photographs. This paper presents a novel technique for categorizing diverse monuments based on the characteristics of their photographs. The deep Convolution Neural Networks VGG16 model is utilized to extract representations. The model is trained on cropped photos of several Indian monuments, which show a wide range of geographic and cultural variety. A monument is a physical structure dedicated to a person, event, or purpose that was built or erected. The importance of this paper to finding and classifying historical monuments accurately without any issue. We used emerging technologies for this identification purpose. without any Machine Learning and Deep Learning, are improving, accelerating image identification development, and allowing computer vision to reach new heights. There is more coverage of international landmarks and monuments, necessitating the need to link a structure's physical presence to its digital presence. As a result, the monument's automated identification comes into action. Almost 100 percent accuracy was predicted using the VGG 16 Model on our proposed dataset.

**Keywords:** Convolution Neural Networks, VGG16 model, Heritage Monuments Identification

1. **INTRODUCTION**

India is one of the world's most beautiful countries, recognized for its vibrant culture and fascinating history. India has a rich and diversified cultural and historical legacy, and its preservation and preservation are critical in today's world [1]. The Rajputana, Dravidian, and Mughal kings created the majority of India's monuments. India's monuments are living memorials to its monarchs' greatness and the genius of its artisans in ancient India [2]. Archaeologists and historians have spent a great deal of time & effort researching many monuments and architectural structures by going to the locations and observing them firsthand. Due to its aesthetic, historical, political, and technical significance, a memorial is a form of building that was specifically built to remember an individual who has become significant to a social cluster as part of the commemoration of historical events [3]. The task of identifying the photographs of monuments based on the architecture is known as Heritage identification of monuments [4]. For every layer of Neural Networks, the following formula is Eq. (1).

g (Wx + b) (1)

Image acquirement

Segmentation

Scaling the Input

Classification

Attribute Extraction

Figure 1: Architectural flow of image classification model [4]

Monument recognition is a difficult challenge in the domain of picture classification. Different structure orientations play a significant part in recognizing monuments in photographs. This paper presents a novel technique for categorizing diverse monuments based on their photographs' characteristics. Deep Convolution Neural Networks are utilized to extract representations. The model is trained on cropped photos of several Indian monuments, which show a wide range of geographic and cultural variety. A monument is a physical structure dedicated to a person, event, or purpose that was built or erected. Because of their artistic, historical, political, technological, or architectural significance, monuments become important to a people as part of their history or culture. Machine Learning and Deep Learning are improving, accelerating image identification development, and allowing computer vision to reach new heights. There is more coverage of international landmarks and monuments, necessitating the need to link a structure's physical presence to its digital presence. As a result, the monument's automated identification comes into action. The goal of this research is to use several Deep Learning architectures to categorize monument photos into their corresponding labels to get the highest accuracy.

The next section discusses the methodology. Section 3 states the results and analysis of monument classification. The final section concludes the paper.

**2. METHODOLOGY**

**2.1 Deep learning**

Deep learning is efficient with a three-layer neural network. It is a subsection of machine learning. These NNs attempt to mimic human brain-like functions by allowing us to learn from massive volumes of data [3]. They include input, hidden, and output layers. The physical activities are done without human interaction. Deep learning knowledge is employed in both current and future products and services [4]. Deep learning means, we use a CNN model. This model works very accurately on image identification and classification.

**2.2 Neural Network**

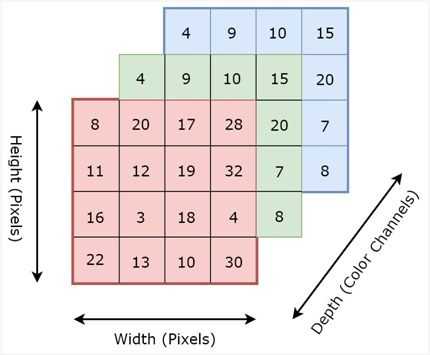
The concept of neural networks comes from biological neural networks found in animal brains. A biological neuron is made up of several organs that contain dendrites, a nucleus, a synapse, and an axon, each of which is responsible for a certain function [6]. In order to represent the synapse stage, we are assigning weights, and for the purpose of simulation of the nucleus, we are using the activation function, and ultimately, the result that mimics axon operation. The biological neuron, as well as the corresponding artificial elements [11].

**2.3 Convolution Neural Network**

A convolutional neural network is a kind of feed-forward synthetic neural network, capable of taking an image as input and assigning importance to the unique perspective or gadgets with inside photos, which differentiates the diverse blanketed units. Influence via the way means of the corporation of the visual cortex, the shape of a convolutional neural network, which is parallel to the animal brain [7]. Assimilating this kind of Structure, a convolutional neural network version is capable of achieving each of the spatial and timing dependencies provided in a photo, imparting the capability to understand even the most challenging images [10]. The CNN can be shown in the following equation [2].

*f* [p, q] \* g [p, q] = (2)

Images may be accompanied by a 2 2-dimensional grid shape, wherein each factor fits a pixel. In an RGB photo, each pixel presents 3 values of intensity, 1 for the 3 number colours, red, green, and blue. Examine that an image has spatial dimensions of one hundred x one hundred, eventually, given the 3 unique colours, the variety of variables could be the same as 100\*100\*3 = 30,000. Immediately, we can expect that for a high-definition image (1080 x 1024), the calculations.



**Figure 2:** 4\*4\*3 RGB Image is extensively enlarges

The major cause of ConvNet is to decrease the pictures into a powerful form, considering the identical time, the sizable elements & functions to preserve the pleasing forecast. In the next figure, we are able to find an RGB photo, with dimensions 4\*4\*3.

**2.4 Pooling layer**

The target of the pooling layer is to lower the scale of the characteristic map, allowing smaller and extra achievable calculations. In other words, the pooling layer simplifies the records contained in the characteristic map and rebuilds it, retaining the maximum beneficial parts. Notwithstanding, max pooling is the maximum well-known one, and exercise appears to paint more effectively. In the figure, we will have a look at an instance of max pooling. The formula for average pooling is shown in Eq. (3).

(*x*) = (3)

Where x represents the values of the given input image segment.

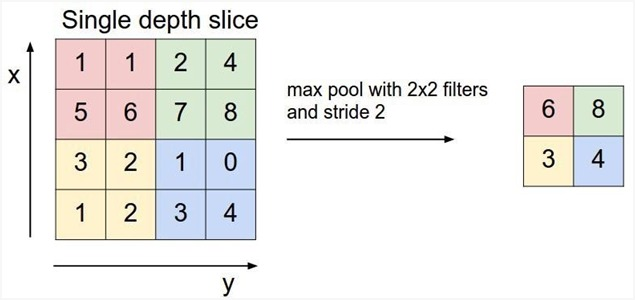


Figure 3:Example of max pooling operation

**2.5 Fully Connected Layer**

Beyond there are numerous varieties of well-known CNN that gift its very own unique structure, and that particular shape will provide splendid outputs. [11]. A few of the maximum famous are Google Net, ResNet, DenseNet, LeNet, AlexNet, and VGGNet. Each version makes use of its very own build, with a one-of-a-kind variety of layers and a unique design**.** A set of dependent nonlinear functions is available in neural networks. A non-linear transformation is applied to the non-linear activation function*f*. It is represented in Eq. (4).

(y) = *f* (4)

**2.6 Mobile Net**

This is a primary computer vision model for mobile applications. This model used depth-wise convolutions. It is dynamically decreasing the number of parameters. Due to this reason, lightweight Deep neural networks are formed [8]. Two operations made depth-wise separable convolution [11].

**3.6.1 Depth-wise Separable Convolution**

The principle behind this convolution is that the depth and spatial dimensions of a sieve can be detached, hence the name separable. Here we used the Sobel filter, which is used to detect edges in images [10]. The depth-wise separation convolution can become a reduction in the computation of Eq. (5):

= (5)

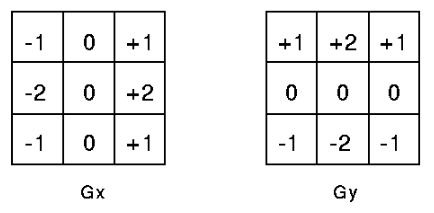


Figure 4: Depth-wise Separable Convolution

The Sobel filter separated height and width dimensions. The Gx sieve is a matrix product of [1 2 1] transposition and [-1 0 1]. We see that the sieve has changed its appearance. It appears to have nine parameters, but it only has six. Because of the separation of its height and breadth parameters, this was achievable. The same principle can be used to distinguish the depth measurement from the plane (width\*height) measurement, resulting in depth-wise divisible convolution. The depth dimension is then covered with a 1\*1 filter. One thing to note is how much this convolution reduces the number of parameters required to generate a similar number of channels.

* The channel-wise DK\*DK spatial convolution is recognized as depth-wise convolution [9].
* Point-wise convolution is the 1×1 convolution to alter the dimensions [10].
* Depth-wise convolution [11].

It's a separate map for each input channel of a single convolution. As a result, the number of outcome channels equals the number of contribution channels. The rate of computing is Df² **\* M \* Dk²**.

* + 1. **Point-wise convolution**

**A kernel iterates over every point of the network.**

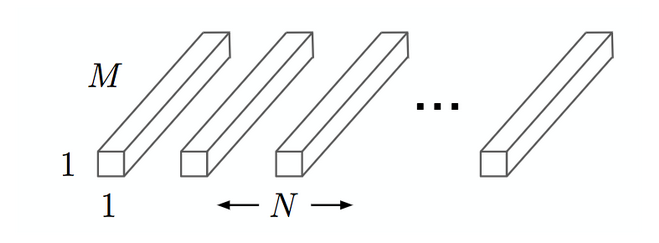
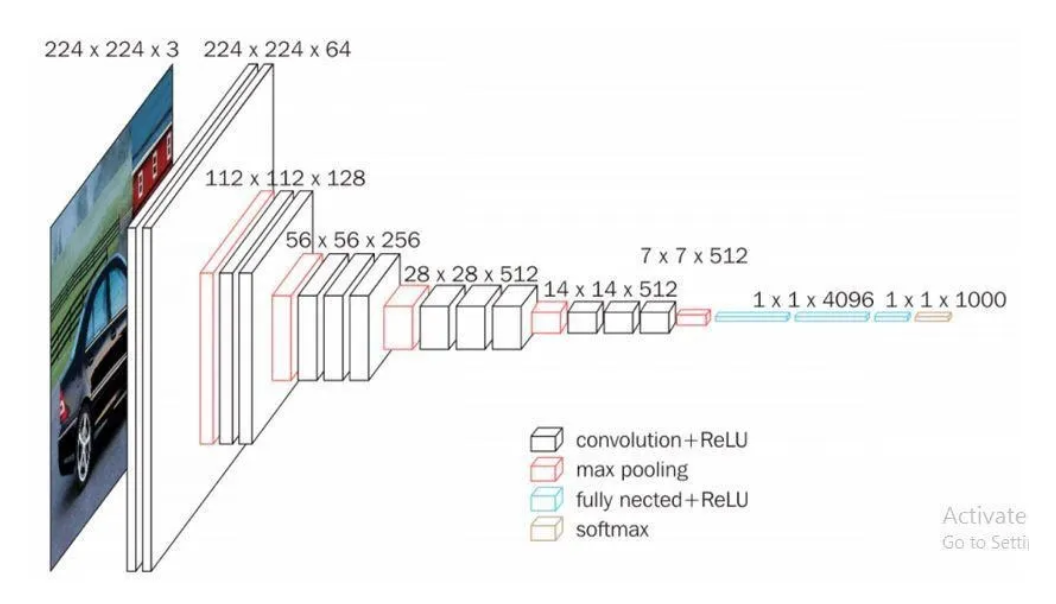


Figure 5: Convolution with a 1x1 kernel scope that merely syndicates the depth-wise convolution's features. M \* N \* Df2 is the computational cost.

**2.7 VGG16 Model**

VGG 16 is a special framework model for CNN with 16 layers. We collect a database from the image for this research. It is divided into 1000 object categories. VGG16 design and clarify the similarities using a use case for object recognition. VGG is a framework to work under CNN for image accurate identification.



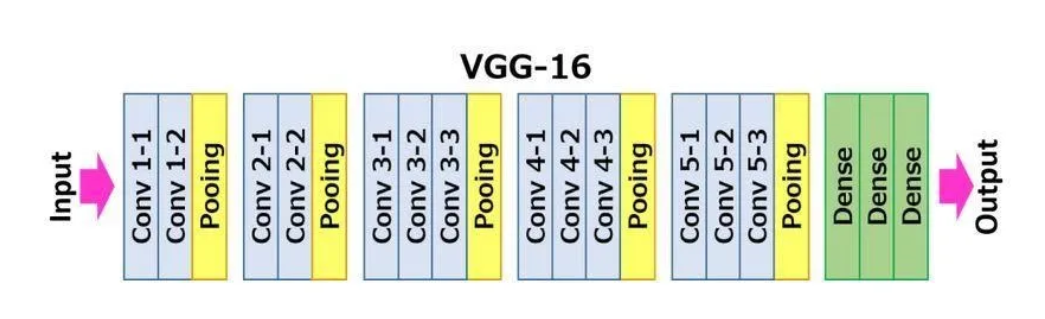


Figure 6: VGG16 Model

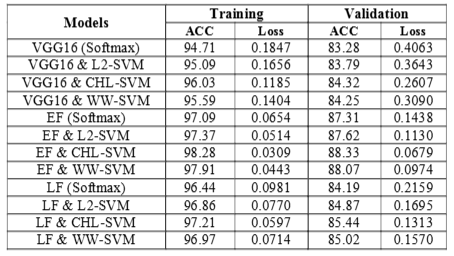
1. **RESULTS AND ANALYSIS**

As a result, the monument's automated identification comes into action. The target of the research is to use several Deep Learning architectures to categorize monument photos into their corresponding labels to get the finest accuracy. The performance of Deep Learning architectures for automated monument prediction is examined and enhanced. Because monuments are three-dimensional structures, the photographs of the monuments have different perceptions. VGG16 is a convolutional neural network model that's used for image recognition.

**4.1 Model Training and Validation Performance**

The following table 1 shows the accuracy and Loss values between training and validation of the model.

**Table 1:** Accuracy and Loss Values Between Training and Validation of the Model



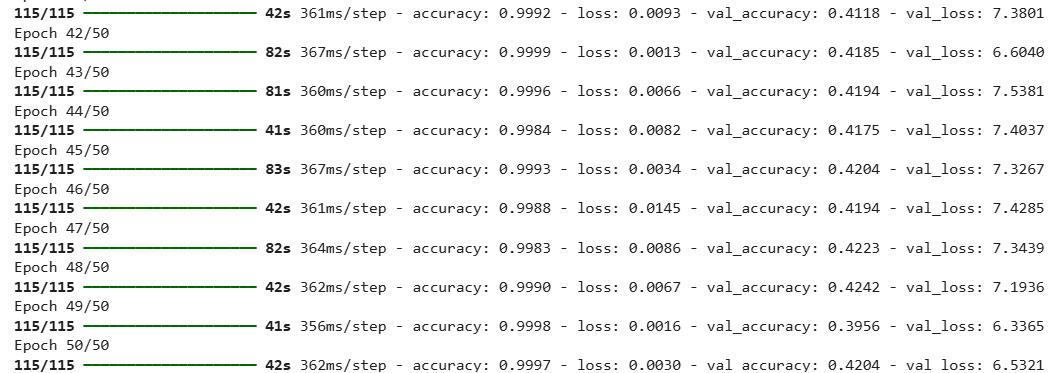
According to Table I, the VGG16 model with dissimilar groupings yielded a training accuracy of 99.28% and a training loss of 0.3122.   
****

Figure 7: Accuracy and Loss Calculation

The implementation of our research to get more than 99% accuracy with less accuracy loss is shown in Figure 7.

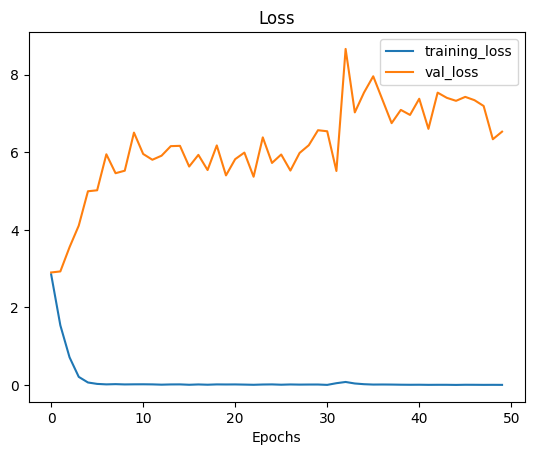


Figure 8: Training Loss with Value Loss

Figure 8 shows the training loss and value loss for monument identification and classification.

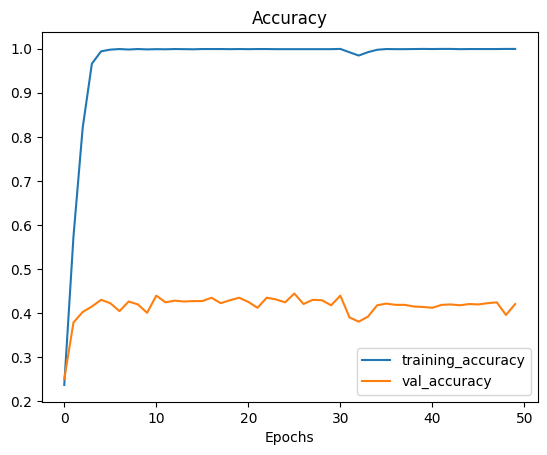
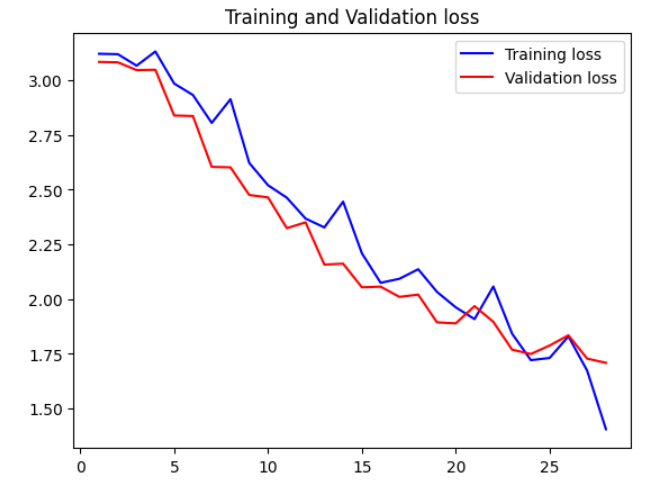


Figure 9: Training Accuracy Rate

The above 9 graphs show the training accuracy and Value Accuracy from the monument classification.

****

**Figure 10:** Training loss and Validation loss

Figure 10 above describes training loss and validation loss. All four graphs display information about accuracy and loss of accuracy with value and validation. The following two historical monuments are predicted using our research implementation accurately without any confusion. Scientifically, it was a big advantage for identifying the historical monuments and classification accurately using emerging technology with the VGG 16 framework.



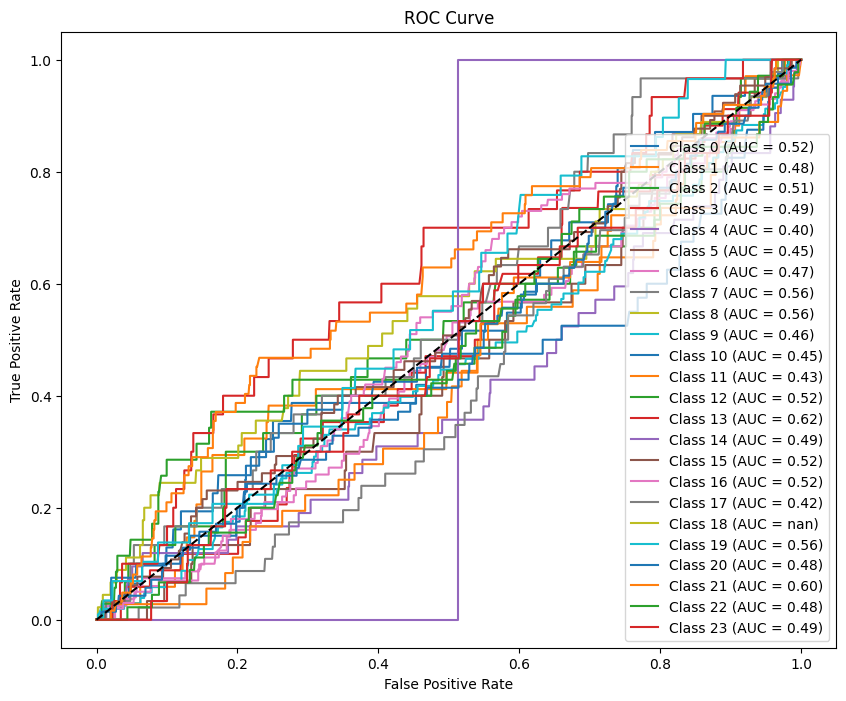


Figure 11: ROC Curve

The above Figure 11 predicts the ROC curve. Receiver operating characteristic (ROC) curves are used to assess the performance of diagnostic tests and machine learning algorithms. They can also be used to compare the performance of multiple tests.

It is an easy perception to realize but useful, especially for humans fascinated by learning about numerous cultures. Our version may be similarly applied in landmark popularity or to broaden an application.

1. **CONCLUSION**

The UNESCO World Heritage Monuments of India, and we have been capable of teaching and following them in real-time item recognition. Make use of the Transfer Learning technique, each fashion is capable of discovering and limiting the preferred items in the company of pleasant precision and high performance. It is an easy perception to realize but useful, especially for humans fascinated by learning about numerous cultures. Our version may be similarly applied in landmark popularity or to broaden an application. In such a manner, a person can have the potential for supplementary statistics with regards to the monument positioned properly in front of him/her, similar to the bodily attributes, past data, gallery with photographs, or maybe a few outermost links. Machine Learning and Deep Learning are improving, accelerating image identification development, and allowing computer vision to reach new heights. As a result, the monument's automated identification comes into action. More than 90% accuracy was predicted using the VGG 16 Model on our proposed dataset.

**REFERENCES**

1. Subhani Shaik,” Analysis of Photo Plethysmography Signals with Artificial Neural Networks Using Curvelet Transform”, Journal of Adv Research in Dynamical & Control Systems, Vol. 11, No. 1, 2019.
2. K. Vaishnavi, G. Pranay Reddy, T. Balaram Reddy, N. Ch. Srimannarayana Iyengar and Subhani Shaik,” Real-time Object Detection Using Deep Learning”, Journal of Advances in Mathematics and Computer Science, Volume 38, Issue 8, Page 24-32, June 2023.
3. Pasumarthi Koushik, Patha Ruchitha, Devarapally Adithya, Subhani Shaik, Sunil Bhutada,” [Deep Learning Approach for Expression-Based Songs Recommendation System](https://link.springer.com/chapter/10.1007/978-3-031-51167-7_56)”, Accelerating discoveries in Datascience and Artificial Intelligence I, ICDSAI 2023, Liet Vizaianagaram, April 24-25, India.
4. Subhani Shaik, “DM Algorithms Based Clustering for Road Accident Data Analysis,” International Journal of Computer Sciences and Engineering, Vol.-6, Issue-9, Sept. 2018.
5. Subhani Shaik,” Analysis of Photo Plethysmography Signals with Artificial Neural Networks Using Curvelet Transform”, Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 1, 2019.
6. KP Surya Teja, Vigneswara Reddy, and Subhani Shaik,” Flight Delay Prediction Using Machine Learning Algorithm XGBoost”, Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 5, 2019.
7. Dr. Vijayalakshmi K andSubhani Shaik, “Predictive Analytics of Employee Attrition using K-Fold Methodologies”, I. J. Mathematical Sciences and Computing, March 2023, 1, 23-36.
8. Md. Samiuddina, G. Rajender, K. Sai Abhiram Varma, A. Ravi Kumar, and Subhani Shaik,” Health Insurance Cost Prediction Using Deep Neural Network”, Asian Journal of Research in Computer Science, Volume 16, Issue 2, Page 46-53, June 2023.
9. K. Vaishnavi, G. Pranay Reddy, T. Balaram Reddy, N. Ch. Srimannarayana Iyengar and Subhani Shaik,” Real-time Object Detection Using Deep Learning”, Journal of Advances in Mathematics and Computer Science, Volume 38, Issue 8, Page 24-32, June 2023.
10. [Eathakota Rupin](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_12#auth-Eathakota-Rupin),  [Tarandeep Singh](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_12#auth-Tarandeep-Singh),  [P. N. Siva Jyothi](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_12#auth-P__N__Siva-Jyothi), [Subhani Shaik](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_12#auth-Subhani-Shaik) &  [Sunil Bhutada](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_12#auth-Sunil-Bhutada),” Office Management Chatbot using Artificial Intelligence”, [International Conference on Soft Computing and Signal Processing](https://link.springer.com/conference/icscsp)(Springer-Scopus), pp 127-142, 16 April 2023.
11. [Mohammad Asif Baba](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_11#auth-Mohammad_Asif-Baba), [Thupakula Sai Varun](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_11" \l "auth-Thupakula_Sai-Varun), [Madasu Shalivahana](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_11" \l "auth-Madasu-Shalivahana), [Subhani Shaik](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_11#auth-Subhani-Shaik) & [N. Ch. Sriman Narayana Iyengar](https://link.springer.com/chapter/10.1007/978-981-99-8628-6_11#auth-N__Ch__Sriman_Narayana-Iyengar)”,  Detection of Abductors using machine learning algorithms”, [International Conference on Soft Computing and Signal Processing](https://link.springer.com/conference/icscsp) (Springer-Scopus), pp 127-142, 16 April 2023.