Alleviate Child Begging by Background-Aware Image Classification

Midhu Jean Joseph
Faculty of Data Science
Mussashino University
3-3-3 Ariake Kotu-Ku, Tokyo, 135-8181, Japan
Email: g2350009@stu.musashino-u.ac.jp

Virach Sornlertlamvanich
Asia AI Institute (AAII)
Faculty of Data Science
Mussashino University
3-3-3 Ariake Kotu-Ku, Tokyo, 135-8181, Japan
Email: virach@musashino-u.ac.jp

Thatsanee Charoenporn
Asia AI Institute (AAII)
Faculty of Data Science
Mussashino University
3-3-3 Ariake Kotu-Ku, Tokyo, 135-8181, Japan
Email: thatsane@musashino-u.ac.jp

Abstract*— *Image classification models has various practical applications ranging from object recognition, facial recognition, sentimental analysis etc. In this work, we plan to utilize the capability of the existing image classification models to identify the child involved in begging, a kind of forced labour in many countries. In many developing nations children are tricked to begging due to circumstantial reasons or forced labour. The first step to address child begging is by identifying the children involved in begging, which currently requires a lot of manual tasks and coordination. Automation of identification of children involved in begging is the focus of the study. Various image classification models of Convolutional Neural Network (CNN) and variations of CNNs (VGG16, EfficientNet) are compared for the usage with the publicly available images for training the model. We propose the use of image background learning to improve the image classification models to detect child begging in public places with 99% accuracy.

Keywords*— *Image Classification, Artificial Intelligence (AI), Convolutional Neural Network (CNN), Child Begging, Transfer Learning, VGG16, EfficientNet, Deep learning models, Image Background Learning

I. Introduction

The expansion of Artificial Intelligence (AI) has taken a big leap. AI is a field of computer science technology that focuses on creating machines models, training with real world data, and making them capable of performing tasks that typically require human intelligence. AI can be used in assisting many difficult and time-consuming tasks to process in a very short span with high accuracy and can be used in addressing many social issues. In many developing nations begging activities are predominantly a social curse and is increasing to become an organized business. An increasing number of children begging on the street (A very common sight in tourist places and traffic light signal in the developing countries) are disturbing and alarming, which need to be recognized for immediate action. At the root level, poverty and social vulnerabilities for example (unemployment, mental health issues, discriminations, lack of education, disability) are identified as the cause of begging and to reduce this government and policy makers must adopt approaches to prioritize social assistance and support welfare programs for the rehabilitation; however, we also need a robust technology assisted system to identify the children forced to begging for taking corrective measures, Here AI comes into rescue. Convolutional Neural Network (CNN), a type of Artificial Neural Network can be used in detecting a begging child from the crowd without human supervision is studied in this paper.

This paper outlines two sets of experiments – (1) To identify the best fit image classification model amongst CNN, VGG16 and EfficientNet. (2) Fine tune the best model (identified from first experiment) with image background learning to achieve higher accuracy. Data set of 600 images are taken from publicly available sources for training the image classification models. The two experiments are designed and conducted with multiple combinations of different number of neural network layers, data set and architecture.

The paper is organized into the following sections: Firstly, with the comprehensive Literature Review and Social Impact to understand the context of societal issues to which application of image classification can be used. Following is the introduction to methodology of the work performed in this study by explaining the experiments and the underlying data. Next section explains the detailed experimentation and the results. The Final sections touches upon the implications and future work.

II. Literature Review and Social Impact

As per Registrar General of India, approximately 45,000 children are involved in act of begging [1]. Child begging also falls under the categorization of child abuse as per WHO, which defines the child maltreatment as those under the age of 18 years of age suffering from physical and/or emotional ill treatment negligence and commercial or other exploitation, which results in actual or potential harm to the child's health, survival, development [2]. Many of the developing countries are facing increased begging activities, at times people/children are forced to take up begging due to different vulnerability factors. The recruitment agents misuse the opportunity as a business. Beggary points to a social disorganization, studies based on Bangladeshi society shows that a significant amount of beggar population are of children, and this has been executed as an organized activity [3].

Children who are forced to begging are at high risk which calls for immediate action. Preventing the child begging needs proper planning and regulations to be in place along with the rehabilitation. For the rehabilitation human actions are very much required, the process of identifying the children in begging can be assisted and well automated using AI supported machine. Deep learning technical solution for child labour detection using CNN and Transfer learning method study had used child labour videos for classification [4]. Keeping Image detection as a base, AI models can be built for detection of child begging using image background

2024 16th International Conference on Knowledge and Smart Technology (KST) | 979-8-3503-7073-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/KST61284.2024.10499679

979-8-3503-7073-7/24/$31.00 ©2024 IEEE

learning by classifying the begging child circumstances into different groups for better training of the model.

This paper discusses the automatic detection of a begging child using the image classification model with image background learning. This model adaptation can be used by communities, institutions, and Government to identify forced begging with help of technology and then streamlining the process of rehabilitation of a begging child, which can bring a major change in the lives of the affected child and the society we live in.

III. Methodology

The technical solution to identify the child involved in begging is suggested by using CNN (Convolutional Neural Network). CNN is a general term refers to a class of Deep Neural Network specifically designed for grid like structured data and uses convolutional layers as base. Convolutional layer performs convolutional operations to detect patterns and features of an image to identify a begging child. Convolutional operations are in simple terms a dot product between two functions, where one function (filter) is a learnable parameter ‘Kernel’, and the other metrics is the restricted portion of the image. The Kernel slides across the image producing an image representation [5]. Two main categories of experiments were performed to Identify a child involved in begging activities. The accuracy (training & validation) were compared to build conclusions on the results.

A. Proposed Experiments and Methods

- Experiment (1): Identifying the best image classification model: Established CNN and Variations of CNNs (VGG16, EfficientNet) are used to train and learn the features of an image to identify a begging child. The best identified model (EfficientNet Functional Architecture) is then used in experiment (2).
- Experiment (2): The best fit image classification model to learn image background techniques: The best CNN model identified (EfficientNet model) is trained to distinguish the background and circumstances (such as traffic signal, crowded markets, etc.). These are the background where the child begging is prominent. This is an approach using classification of the images background of a begging child for better training of the model for better accuracy.

B. Underlying Data

- *A* unique dataset of images of Begging child and Normal child is created using keyword search in google search bar like “Begging child India”, “Normal child India”, “Begging child Mumbai”, “Begging child Delhi”, “Begging child Chennai”, “child in school India”, “Normal child in street India”, “children in play area India”, “normal children in traffic areas India” etc.
- All the images were shortlisted from India specifically from appearing in online published newspapers, magazines, forum etc.
- Data set of 600 images (300 Begging child and 300 Normal child was created); however different sample sizes of images were used in each experiment (Details are given in respective columns “Number of Training samples “ and “Number of Validation samples” in Table I to V) and classes were equally distributed.
- In the experiments model architecture, the image was resized to 150*150 pixel for CNN and 224*224 pixel for VGG16 and EfficientNet.

IV. Experiments and Results

A. Experiment (1): Identification of best image classificaiton model for the identification of the begging child.

Existing 3 established image classification models with architectural variations were trained and tested under this experiment – a) CNN, b) VGG16 c) EfficientNet.

a) Experiment (1a)_Basic CNN Model: CNN can have multiple Layers of Convolutional layer (Conv), Pooling Layer (PL), Fully Connected layer (FC) and flexibility of having different numbers of these layers and different filter size (Kernel). Three different sub-experiments with multiple combinations of Conv, PL, FC, and Batch Normalization(BN) were used for training and testing to choose the best CNN model (Details in Table I)

- CNN_1: Initialized the experiment with 3 Conv, 3 PL, 1 FC layers and 200 image sample size.
- CNN_2: Increased by an additional layer of Conv, PL, and FC along with Batch Normalization to improve feature extraction and accelerate training.
- CNN_3: Increased the Conv layer to 8 and the training image size increased to 500 for improving generalization of model with more diverse examples.

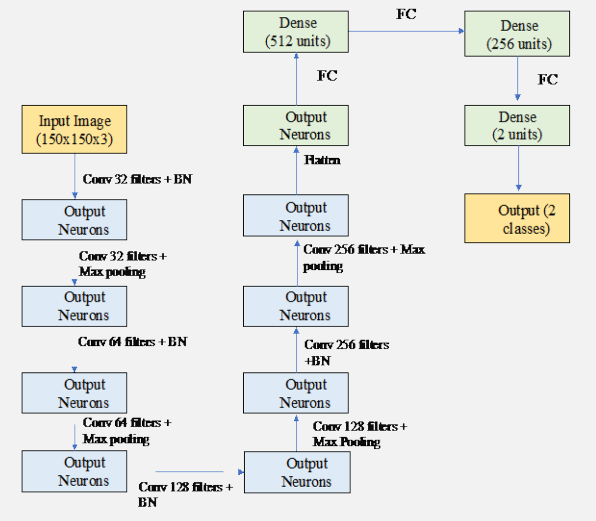

Fig. 1. CNN architecture (Experiment CNN_3)

- <u>Observations from CNN model experiments:</u> Increasing conv layers, FC layers and the size of training data improved validation accuracy and reduced validation loss; however, these models showed overfitting (Overfitting: Training accuracy > Validation accuracy; Validation loss: Discrepancy between the prediction and the true value). CNN experiments gave validation accuracy in the range of

70% -77% (Accuracy details are in Table I and Figures 8 to 10).

b) Experiment (1b)_ VGG16 Model: VGG16 has specific CNN architecture with 16 layers (13 conv , 3 FC layers & a fixed 3x3 convolutional filter). This model captures complex patterns, fine grained features for image identification (Details in Table II)

- VGG16 Base: The base model of 13 Conv and 3 FC layers was used to initialize the experiment.
- VGG16 Transfer 1 (CNN): VGG16 pre trained model serves as base model, by leveraging the pre- trained model and adding custom classifier layers transfer learning improves performance.
- VGG16 Transfer 2 (CNN): Like the previous transfer learning experiment but with reduced training image size to study and compare the outcome.

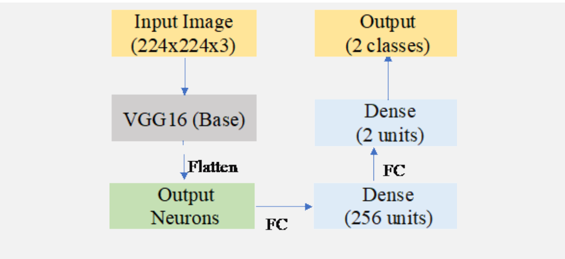

Fig. 2. VGG16 architecture (Transfer Learning)

- Observations from VGG16 model experiments: All VGG16 models achieved 99% training accuracy and 85% validation accuracy (after 25 epochs run); However, overfitting continues even for VGG16. Transfer Learning with reduced training samples showed reduced validation loss (Accuracy details in Table II and Figures 11 to 13)

c) Experiment (1c)_EfficientNet Model: EfficientNet optimize the model's architecture in a principled manner by scaling the depth, width and resolution simultaneously, this model is more efficient in terms of memory and computational requirements. There are two architectures tested: Sequential and Functional architecture (Details in Table III)

- EfficientNet (sequential): The experiment with sequential architecture is using most simple linear stack of layers.
- EfficientNet_1 (Functional): The Functional architecture used in this model is more complex allowing more flexibility to the model.
- EfficientNet_2 (Functional): Increased the FC layer in the functional architecture to find the difference in the outcome.

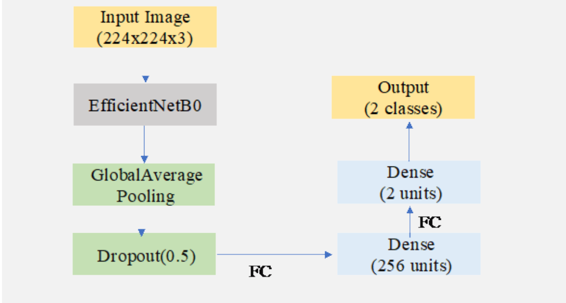

Fig. 3. EfficientNet_2 Functional architecture

- Observations from EfficientNet model experiments: The EfficientNet model with Sequential architecture removed the overfitting issue with an accuracy at 50%; However, the EfficientNet model with functional architecture and additional FC layer achieved 98% training and 99% Validation accuracy and this model shows underfitting (Training accuracy < Validation accuracy) (Accuracy details in Table III and Figures 14 to 16)

Result of Experiment (1): CNN, VGG16 and EfficientNet models can be used for detections of child begging. In Experiment (1) EfficientNet model (Functional) gave the best results. Hence chose this model to upskill by image background learning in further experiments.

B. Experiment (2): Training best fit Model to learn Background and Foregrounds to enhance the accuracy

For image background learning, the technique used was to divide the images into 14 subcategories based on the location/surroundings where the child is found begging and learn the features on those classification.

- Begging: Market Areas, Public places, Traffic areas, Street and over bridge
- Normal: Market Areas, Public places, Traffic areas, Street, over bridge, ghettos, House, School, and Kids play areas.

The above 14 sub-categories were used to train the model in 2 different methods. The model EfficientNet-Functional architecture which gave the best result in Experiment (1) was used here for the image background learning method.

a) Experiment (2a)_Learn with Sub-categories: The images are categorized into 14 sub categories based on the various image background within Begging and Normal circumstances (Figure 4). Results are shown in Table IV).

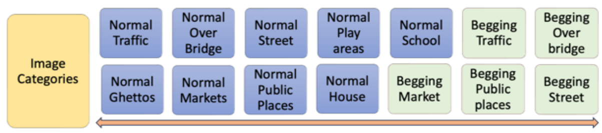

Fig. 4. The image sub-categories structure as per the background

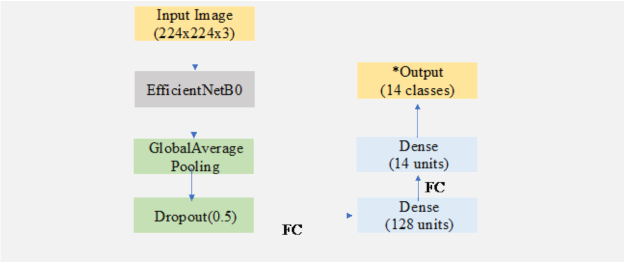

Fig. 5. EfficientNet Learn with Sub- categories architecture.
**The highest probability amongst subcategories is used to determine begging or normal child*

b) Experiment (2b)_Learn with Subcategories and Sub-labelling: In the second method, one extra layer for learning the image background is added by sub-labelling the images to Begging and Normal and further sub-categorizing as per the 14 backgrounds (Figure 6).

- Observations from subcategories and sub-labelling model experiment: The models achieved 98% training accuracy and the validation accuracy reached 99% (after 100 epochs). There was less underfitting observed. (Accuracy details in Table V and Figure 18)

Result of Experiment (2): The Experiment (2) model architecture utilized image background learning through subcategories and sub-labelling resulting in improving the overall accuracy.

- Observations from sub-categories model experiment: The model achieved 85% training and 97% validation accuracy (100 epochs) and underfitting is seen in this model (Training accuracy < Validation accuracy). (Accuracy details in Table IV and Figure 17)

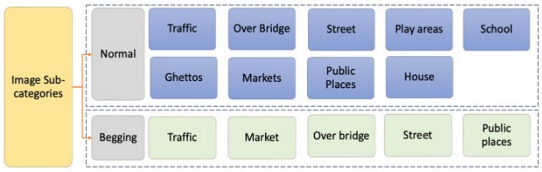

Fig. 6. The Image subcategories sub-labelling background structure

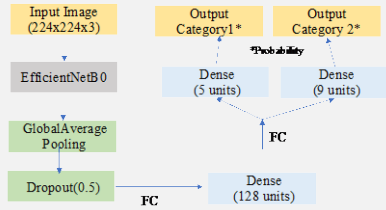

Fig. 7. EfficientNet Learn with Sub- categories architecture.
**Category 1 and Category 2 refers to two sub-categories (Begging or Normal)*

TABLE I. CNN Experiments and their performances

S.No	Number of Conv Layers	Number of Pooling Layers	Number of FC Layers	Number of BN Layers	Number of Training samples	Number Validation samples	Training Accuracy	Validation Accuracy	Reference
CNN_1	3	3	1	0	100	100	99%	70%	Fig. 8
CNN_2	4	4	2	4	100	100	99%	70%	Fig. 9
CNN_3	8	4	3	4	500	100	99%	77%	Fig. 10

TABLE II. VGG16 Experiments and their performances

S.No	Number of Conv Layers	Number of FC Layers	Number of Training samples	Number Validation samples	Training Accuracy	Validation Accuracy	Reference
VGG16 Base	13	3	500	100	99%	86%	Fig. 11
VGG16_Transfer1_ (CNN)	13	2	500	100	99%	85%	Fig. 12
VGG16_Transfer2_ (CNN)	13	2	100	100	99%	85%	Fig. 13

TABLE III. EfficientNet experiments and their performances

S.No	Number of Conv Layer	Number of FC Layer	Extra Number of FC Layer	Number of Training samples	Number Validation samples	Training Accuracy	Validation Accuracy	Reference
EfficientNet (sequential)	20	1	1	500	100	50%	50%	Fig. 14
EfficientNet_1 (Functional)	20	1	0	500	100	96%	90%	Fig. 15
EfficientNet_2 (Functional)	20	1	1	500	100	98%	99%	Fig. 16

TABLE IV. Background Learning EfficientNet with Subcategories experiments and their performances

S.No	Number of Conv Layer	Number of FC Layer	Extra Number of FC Layer	Number of Training samples	Number Validation samples	Number of epochs	Training Accuracy	Validation Accuracy	Reference
EfficientNet Learn with Sub-categories	20	1	1	500	100	100	85%	97%	Fig. 17

TABLE V. Background Learning EfficientNet with Subcategories and Sub-labelling experiments and their performances

S.No	Number of Conv Layer	Number of FC Layer	Extra Number of FC Layer	Number of Training samples	Number Validation samples	Number of epochs	Training Accuracy	Validation Accuracy	Reference
EfficientNet Learn with Subcategories and Sub-labelling	20	1	1	500	100	200	98%	99%	Fig. 18

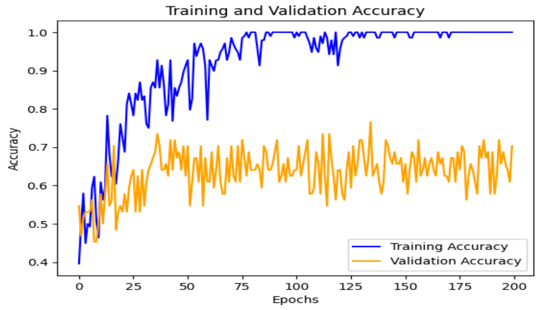

Fig. 8. Training and Validation accuracy Vs epochs (CNN_1)

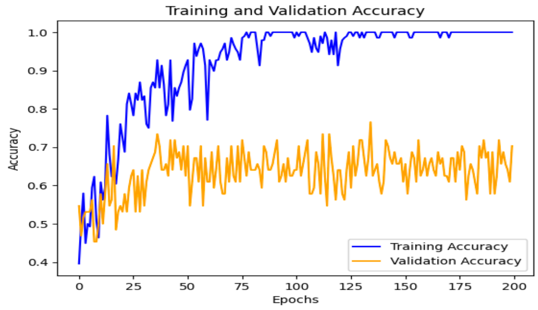

Fig. 9. Training and Validation accuracy Vs epochs (CNN_2)

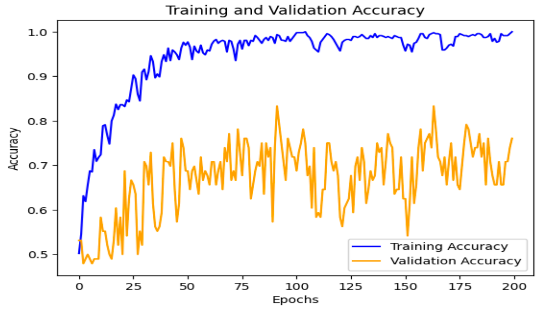

Fig. 10. Training and Validation accuracy Vs epochs (CNN_3)

Fig. 11. Training and Validation accuracy Vs epochs (VGG16 Base)

Fig. 12. Training and Validation accuracy Vs epochs (VGG16_Transfer1)

Fig. 13. Training and Validation accuracy Vs epochs (VGG16_Transfer2)

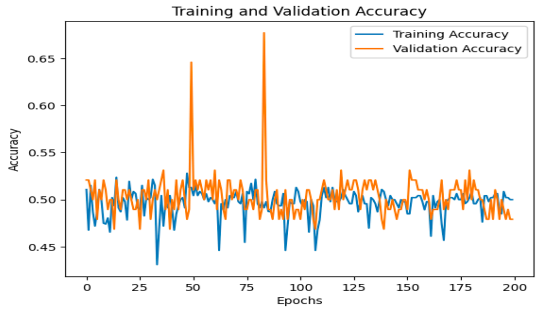

Fig. 14. Training and Validation accuracy Vs epochs (EfficientNet sequential)

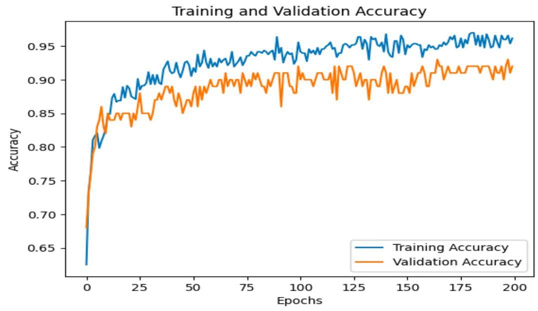

Fig. 15. Training and Validation accuracy Vs epochs (EfficientNet1 Functional)

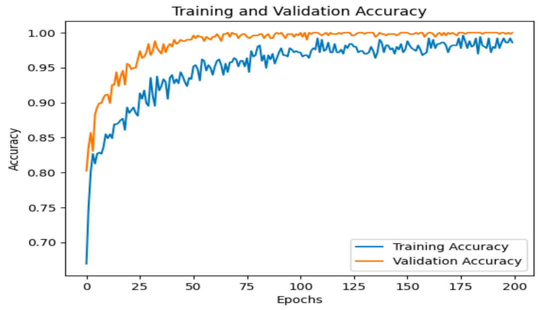

Fig. 16. Training and Validation accuracy Vs epochs (EfficientNet2 Functional)

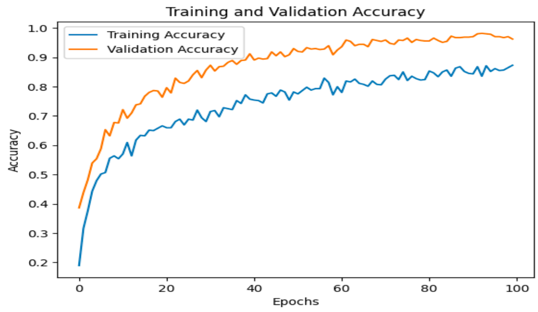

Fig. 17. Training and Validation accuracy Vs epochs (Learn with Subcategories)

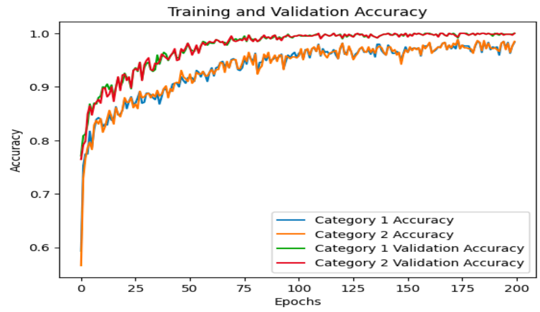

Fig. 18. Training and Validation accuracy Vs epochs (Learn with Subcategories and Sub-labelling

**Category 1 and Category 2 refers to two sub-categories (Begging or Not begging)*

V. IMPLICATIONS & FUTURE WORK

Neural Network model with image background learning shows that a robust AI assisted system can be built for identification of the child involved in begging which can aid for the rehabilitation process. The AI assisted systems (AI Cameras) can be utilized in those areas where these activities are more common: traffic signals, crowded tourists' spots, busy markets. The details obtained from these AI systems can be utilized by government, NGOs, etc. for actions and policies to reduce the child begging. In Future, following can be explored as an extension of this study:

1) Identifying better model in terms of efficiency (accuracy) and economically viability (less computational expenses, easy to implement).

2) Improving current models with improved learning on image backgrounds, emotions depicted, other feature mappings like posture detection which will be an added advantage to the current model.

REFERENCES

[1] Government of India, Press Information Bureau Delhi, "Child Begging", [Press release]. Retrieved from https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1778853, , 7 December 2021.

[2] World Health Organization, "Child maltreatment", Retrieved from https://www.who.int/news-room/fact-sheets/detail/child-maltreatment, 9 September 2022.

[3] Al Helal, M. A., & Kabir, K. S, "Exploring Cruel Business of Begging: The Case of Bangladesh", Asian Journal of Business and Economics, Volume 3 (3.1), ISSN: 2231-3699, 2013.

[4] Progga, F. T., Shahria, M. T., Arisha, A., & Shanto, M. U. A, "A Deep Learning Based Approach to Child Labour Detection", 2020 6th Information Technology International Seminar (ITIS). DOI: 10.1109/ITIS50118.2020.9321027, October 2020.

[5] Mishra Mayank, "Convolutional Neural Networks, Explained.", Towards Data Science, 27 August 2020.

[6] Yu, Y., & Liu, W., "Ethical Issues of Child Abuse-A Cultural Comparison." In 2020 International Conference on Public Health and Data Science (ICPHDS), IEEE. DOI: 10.1109/ICPHDS51617.2020.00017, 2020.

[7] Khan, J. H., Menka, Dr., & Shamshad, "Socio-Economic Causes of Begging", International Research Journal of Human Resources and Social Sciences, Volume-1, Issue-3, ISSN: 2349-4085 Page 37, August 2014.

[8] Ayoob, SM, "Beggary in the Society: A Sociological Study in the Selected Villages in Sri Lanka", Journal of Xi'an University of Architecture & Technology, Volume XI(Issue XII), Page 1725, 2019.

[9] Kai Yuanqing Xiao, Logan Engstrom, Andrew Ilyas, Aleksander Madry. "Noise or Signal: The Role of Image Backgrounds in Object Recognition", Published as a conference paper at ICLR 2021.

[10] Dheeraj Mekala, Keerti Anand, Aakash Ghosh,Saurabh Kataria,. "Foreground-background classification and ROI detection in surveillance videos", 2016.

[11] Mazda Moayeri, Phillip Pope, Yogesh Balaji, Soheil Feizi. "A Comprehensive Study of Image Classification Model Sensitivity to Foregrounds, Backgrounds, and Visual Attributes". IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022