Journal of Information Hiding and Multimedia Signal Processing ©2023 ISSN 2073-4212
Ubiquitous International Volume **14**, Number **3**, September **2023**

Tribal Dress Identification using Convolutional Neural Network

Md. Fazle Rabbi, Md. Nahid Sultan, Mahmudul Hasan, Md. Zahidul Islam

Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University
Dinajpur -5200, Bangladesh
rabbi@hstu.ac.bd, nahid.sultan@hstu.ac.bd, mahmudulmoon123@gmail.com, zahidulislam.cs@gmail.com

Received May 2023; revised August 2023

Abstract. *Tribal people have a unique identity in their culture, food habit, and dress. People are getting creative with their style by fusing elements of indigenous attire with contemporary trends. People are getting interested in tribal dress, and online shopping plays a vital role in this segment. Due to a lack of knowledge to identify the tribal dress, people are getting fraud online. This study developed a Convolutional Neural Network (CNN) based model that can identify the tribal dress by capturing the pattern of the dress. We compare the performance of CNN with the performance of VGG16 and VGG19 to find the best model for classification. A tribal dress dataset is created by us and named Tribal Dress Dataset that contain the dress of two popular tribal Chakma and Monipuri. Data augmentation using Generative Adversarial Network (GAN) enriches the dataset to make it suitable for training the deep learning models. Accuracy, Precision, Recall, F-1 score, and ROC curve are used to measure the performance of CNN. CNN shows 97.30% accuracy, 97% precision, recall, and F1 score on the augmented dataset. Integrating this model is online marketplace can help the customers to ensure the quality and originality of the tribal dress.*

Keywords: Tribal dress identification, tribal dress dataset, deep learning, Convolutional Neural Network.

1. **Introduction.** Dress is considered one of the important symbols of the identity of people in society. The choice of fashion is changing according to changing the times. Nowadays, people try to wear something unique and different. Fashion designers are mixing the traditional concept with modern fashion to create something new [1]. In Bangladesh, the tribal people have a unique culture. Most tribes have their language, dress, food habits, rules, and regulations. The dress is important in uniquely identifying one tribe from another [2]. They make their dress by hand, which is comfortable to wear. Tribal dresses are getting famous day by day to us. Many fashion designers are mixing the traditional tribal dress with exemplary fashion and creating amazing fusion dresses. Also, the tribal dress is now easily be found for online shopping. This Covid-19 pandemic has changed many things and made many entrepreneurs in different sectors. But the online shopping has some problems that the attention of the people is getting down for the activities [3]. People are getting low-quality products, some unknown products like the tribal dress, and foreign products that are unknown to them. Most people do not know the pattern of different tribal dresses, and most see the dress for the first time. They are getting confused and fraud by the sellers for lack of knowledge of tribal dress identification.

The different tribal dress has different patterns [4]. Most dresses are followed by a pattern, which will be the feature to identify the dress uniquely [5]. Much research is already done to identify dress, types of dress, and patterns of dress, but identifying the tribal dress are not made previously [6] [7]. Data collection, processing, and then training and testing are performed to identify the dress of the tribal. For tribal dress identification, the contributions of this paper are:

- We proposed a deep learning based setup that can classify tribal dress more accurately.
- We create a dataset that contain the dress of two popular tribal in Bangladesh.
- We successfully enrich the dataset by applying GAN as image augmentation techniques.
- We compair some neural network model to find the suitable model for triabal dress classification.

Here is how the paper is structured: In Section 2 covers some similar studies in this subject, while Section 3 presents the proposed dataset development. Data preprocessing techniques, a brief overview of the algorithms, and several preperformance measure techniques are described in Section 4. The findings and discussion are presented in Section 5, while a summary and outlook are provided in Section 6.

2. **Literature Review.** Already many works are done by researchers to classify dress using different classification algorithms. But, we do not find any specific paper on tribal dress classification. Some dress identification and classification-related works are below. In a study [8], researchers propose to apply Hierarchically Convolution Neural Networks (H-CNN) on apparel classification. VGGNET implements it, and the fashion MNIST dataset is considered in this case. Using 128 batch size, it achieves 93.33% accuracy in sixty epochs at a 0.001 learning rate. It shows better performance than the base model VGG19. But this study does not consider the hierarchical structure of the dataset in the case of classification.

In another study by Tejaswini et al. [9], authors applied machine learning techniques to identify different dress characteristics in several phases. Firstly, the target localization is done by a segmentation algorithm named SegNet with the LIP dataset. After that, human key joints are identified by a model developed by researchers at Carnegie Mellon University, and a bounding box is generated over the dress region. Finally, their proposed method classifies the dress' hem length, style of dress sleeves, and hem style of dress and compares the result with CNN, VGG19, and VGG16. In the hem style classification, their model gets the greatest score of 96.7%, whereas CNN, VGG16, and VGG19 have 86.4%, 73.3%, and 74.8%, respectively. Another Novel approach to clothing classification is worthy of remark. Researchers of this study [10] attempted to classify clothes based on physical characteristics and selection masks. They attempted 3 approaches with 5-fold cross-validation, and among these approaches, their LHCS approach topped, scoring 90% in average TPR. Jan et al. presented a set of experiments in their study [11] that established Convolutional neural networks like ResNet, Squeezenet and SSD. Their procedure includes detecting dresses in the image with SSD300, augmenting a natural background for studio image and final classification of the dress with five other attributes like Color. The classification was done in 4 ways, single squeeze net, single resnet, ensemble resnet, and ensemble squeezeout and among them, the latter produced the best accuracy of 78, 75, 90, 77, and 88 for Color, Pattern, Sleeve, Neckline, Hemline, respectively. Kowshik et al. [12] attempted to automate the process of dress code checking in their research by leveraging a three-layer Convolution Neural Network. They Collected 270 pictures of people with proper and improper dress codes. After their experiments, they achieved more than 70% in all the constraints.

3. **Dataset.** The dataset used in this analysis is created by collecting the images from different sources. Chakma and Monipuri, two tribal dress images, are collected, and most images are captured from their shop in Rangamati and Sylhet. In the original dataset, there are 150 images in Chakma and 120 images in Monipuri segment. A dataset sample is in Fig. 1. The different tribal dresses contain patterns representing their unique identity. Most of the designs are created by tribal women, and the patterns are different from one tribe to another. The image Augmentation technique enriches the dataset to fit a deep learning model. The augmentation techniques are described in the next section.

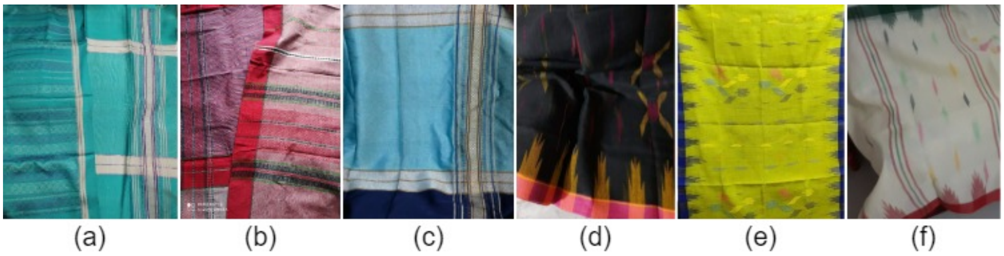

FIGURE 1. Example of the tribal dress images. (a-c) represent Chakma dress and (d-f) represent Monipuri dress.

4. **Proposed Mechanism.**

4.1. **Overview of Proposed Methodology.** The identification model is constructed using CNN, VGG16 and VGG19. Data collection, image processing, model training, testing and performance evaluation are the step-by-step process of the proposed methodology. Fig. 2 shows an overview of the proposed methodology. The description of CNN, VGG16, and VGG and other mechanisms are also described in the next subsection.

4.2. **Data preprocessing.** The images of the dataset are captured by camera manually. We crop the necessary portion of each image to make it more trainable for the models. We also store the images to individual folder as per the label of the image.

4.3. **Image Augmentation Technique:** An exciting new method for improving and expanding datasets used to train machine learning models, especially for computer vision applications, is image augmentation using GANs (Generative Adversarial Networks). Increasing the variety of training data is a common practice, and traditional image augmentation techniques like rotations, flips, and color modifications are frequently employed for this purpose. More than that, though, GAN-based augmentation can create realistic and diversified images that weren't included in the original dataset. It's important to note that training a GAN can be resource intensive on the computer and that steady training necessitates fine-tuning. The process of augmenting images can be made easier by using pre-trained GAN models designed for certain applications. There may have been new developments and advances in this area since my last report in September 2021, when GAN technology was still in its infancy. Then image GAN augmentation is used to enrich the dataset for the model. Augmentation is done with $ImageDataGenerator$ class of $keras$. For each image with have 30 augmented image with the following modifications: width shift range(-20,20), height shift range(-20,20), horizontal flip, rotational range(20), rescale(1/255), shear range(0.1), zoom range(0.10), brightness range(0.2, 1.0), fill mode(reflect).

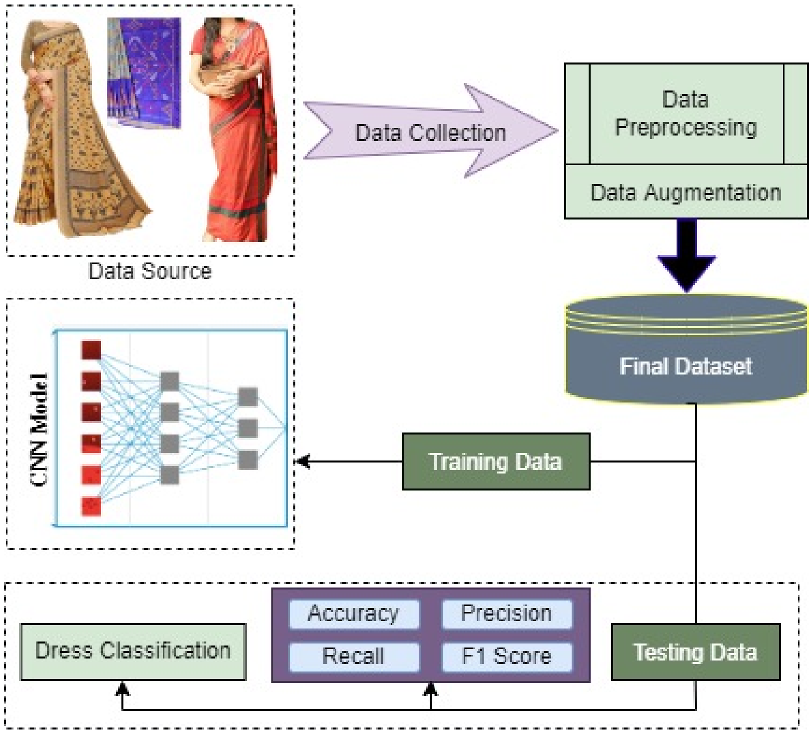

FIGURE 2. Overview of proposed dress identification system

4.4. **Description of the Models. CNN:** In this analysis, we use the CNN deep neural network to classify the tribal dress. A short description of the CNN algorithm is described below. CNN (From Flood Paper) is a deep learning algorithm used for feature extraction, generating any pattern mapped from input to output and others. It is one kind of feed-forward neural network. Time series datasets, computer vision, image processing, speech recognition, and so on are widely used, providing reasonable performance [13]. CNN has 6 layers. They are: Input layer, convolutional layer, pooling layer, fully connected layer, softmax layer, output layer.
(I) Input layer: It consists of input data and can be represented using a matrix.
(II) Convolutional layer: The convolutional layer is responsible for feature extraction and mapping activity. Let's assume the time series data of flood is of the form P = (P0, P1, P2.........Pn). So, the kth convolutional layer is defined as this form.
(III) Pooling layer: After convolution, this layer is applied to decrease the spatial volume of data streams for the next layer. It will be computationally costly to deploy a fully connected layer after the convolution layer without using pooling, which we avoid by using a pooling layer. It's utilized inside two convolution layers. Average pooling and max pooling are two deviations from it [14].
(IV) Fully connected layer: CNNs, which have shown to be particularly useful in image recognition and classification for computer vision, require fully connected layers. Convolution and pooling are used to break down the image into features, which are then analysed individually. This method produces a fully linked neural network structure, which drives the final classification decision.
(V) Softmax layer: Softmax is used to solve classification problems. Before using softmax, some vector components may be negative or greater than one, and they may not add to one. The softmax layer outputs a probability distribution, and the values of the output sum to one. This added constraint allows training to converge faster than it would without.

(VI) Output layer: The CNN's final layer is represented by the output layer. It is made up of the same number of neurons as the number of unique classes in the classification task. The output layer offers classification probabilities or predictions for each class, indicating the likelihood of the input image belonging to a specific class.

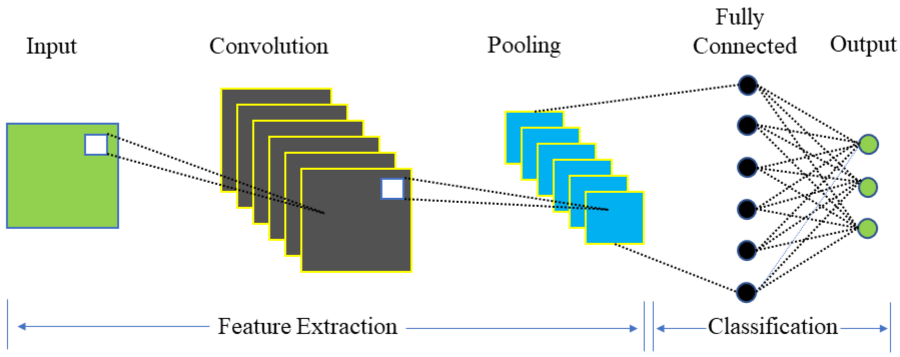

FIGURE 3. Block diagram of CNN

VGG16: The Visual Geometry Group at Oxford University created the VGG16 deep convolutional neural network architecture in 2014 [15]. It's well-known for being both easy to use and effective in image identification tasks, because to its 16 layers, 13 of which are convolutional layers, and 3 of which are completely connected.

Convolutional filters of size 3x3 and stride 1 are used in VGG16 and are stacked to build more accurate approximations of the input image [16]. By using these simple filters, the network may be trained to recognise increasingly complex features with a smaller set of inputs.

In several computer vision tasks, including object identification, image segmentation, and visual question answering, VGG16 has been utilised as a feature extractor and as a basis model for transfer learning.

VGG19: Similar to VGG16, but with 19 layers instead of 16, VGG19 is a deep convolutional neural network architecture [17]. The Visual Geometry Group at Oxford University created it in 2014, and it has since become widely used due to its superior accuracy in image categorization tasks.

Convolutional layers are followed by completely connected layers in VGG19, just as they were in VGG16. VGG19 differs from VGG16 in that it contains four extra convolutional layers, allowing it to pick up more nuanced features during training [18].

To further downsample the feature maps, VGG19 employs small convolutional filters (3x3) with a stride of 1 and max pooling layers. The network is able to learn more discriminative features while the feature maps remain manageable in size because to this design decision.

VGG19, like its predecessor VGG16, has seen extensive use as a foundational model for transfer learning and feature extraction across a wide range of computer vision applications. On multiple state-of-the-art datasets, including ImageNet, CIFAR-10, and CIFAR-100, it has achieved state-of-the-art performance.

4.5. **Performance measure techniques.** To determine the performance of ML and DL algorithms, this study employed five performance measurement techniques as Accuracy, Precision, Recall, F-1 Score and ROC.

Accuracy: Data classification accuracy is defined as the proportion of instances that were correctly labelled by the classification algorithm. One of the most fundamental measurements of performance, accuracy can be misleading for unbalanced datasets [19].

Mathematically,

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

Precision: The ratio of true positives (TP) to false positives (FP) is the definition of precision in binary classification. When the goal is to lower FP, Precision functions flawlessly on skewed data. A high FP rate need not preclude the use of accuracy as a metric. Mathematically,

$$Precision = \frac{(TP)}{(TP + FP)} \tag{2}$$

Recall: True Positive Rate (TPR) or Sensitivity are other names for recall. As a rule of thumb, recall equals the ratio of true positives to true positives plus false negatives. Recall is appropriate for FN reduction from the unbalanced dataset [20]. Mathematically,

$$Recall = \frac{(TP)}{(TP + FN)} \tag{3}$$

F-1 Score: F-1 score is the mathematical mean of Precision and Recall. Model applicability cannot be determined by accuracy alone. Precision and Recall must be high for the model to make any sense. This is why the F1-Score is used to evaluate the efficacy of different classifiers. Higher f-1 scores indicate a more reasonable model, with a range of [0, 1]. Mathematically,

$$F1\_Score = \frac{(2PR)}{(P + R)} \tag{4}$$

Receiver Operating Characteristics (ROC): The ROC curve is a method for comparing the relative efficacy of different classifiers. It's a graph showing how the false positive rate (FPR) changes when the threshold level (TPR) changes.

5. **Results and Discussion.** CNN classifies the dress of the tribal. Using the proposed methodology, we classify the dress of two different tribals of Bangladesh. CNN shows good accuracy with good precision, recall and F1 score also. The confusion matrix is shown in Fig. 4. The Figure clearly indicates the supriority of CNN in classification. The true negative and true positive rate is extremely high and the flase positive and true negative rate is too low. It indicates the high precision and recall of the CNN. It also indicates that the rate of misclassification of CNN is too much low in this setup. We only considered Chakma and Monipuri dresses for the availability of the data. CNN shows one false negative and zero false positives, proving it performs well.

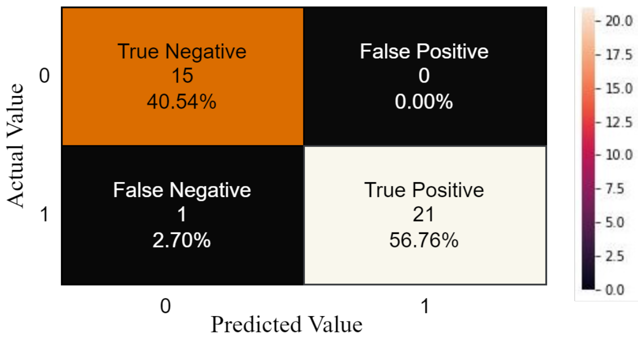

FIGURE 4. Confusion matrix of CNN to classify tribal dress

Table 1 shows the accuracy with precision, recall, and F1 score of the CNN. It shows 97.30% accuracy and 97% precision, recall and F1 score. It truly identifies the two tribal dresses correctly. In Fig. 5 the accuracy of training and validation is shown. In the beginning, the curve shows the difference in the training and validation accuracy. At the last epochs, the performance is good than the beginning. The performance of VGG16 is 93.50% with 92% precision, 91% recall and 92% f1 score. The VGG19 shows 90% precision, recall, f1 score and accuracy. CNN show its superiority than other two models.

Table 1. Performance evaluation on tribal dress classification

Algorithms	Precision	Recall	F1	Accuracy
CNN	0.97	0.97	0.97	97.30%
VGG16	0.92	0.91	0.92	93.50%
VGG19	0.90	0.90	0.90	90.00%

Also, Fig. 6 shows the training and validation loss. At the beginning epoch, the loss difference is high, but the loss decreases with the increase of the epochs. Due to a lack of data, the curve is not smooth from the beginning, and sometimes the accuracy goes up and down. Also, images were collected using different devices, and the capturing was not systematic.

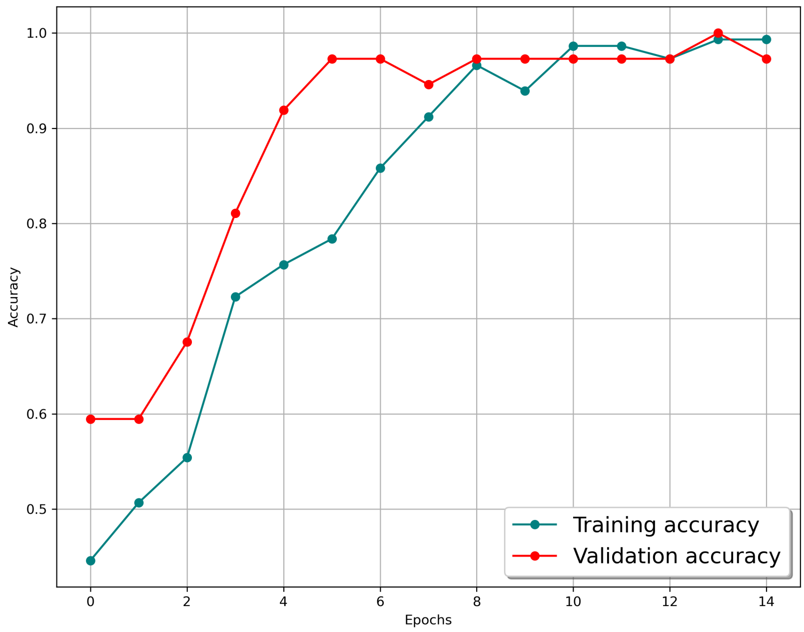

Figure 5. Training vs validation accuracy curve of CNN.

Consumers have a wide variety of conceptualizations of what constitutes various "fashions" or "classifications" in the fashion industry and online shopping. Lack of knowledge, customers are getting fraud into buying tribal dresses. Our proposed model can identify the dress more accurately, and customers are getting benefited from this model.

6. **Conclusion and Future Work.** This study uses CNN to classify the tribal dress from a dataset we generated. Only two tribal dresses are taken in the dataset. Using image augmentation, new images are generated to make the dataset rich. CNN shows a satisfactory result for identification. Using this model, we can easily identify the tribal dress that helps us to buy the original dress online or offline.

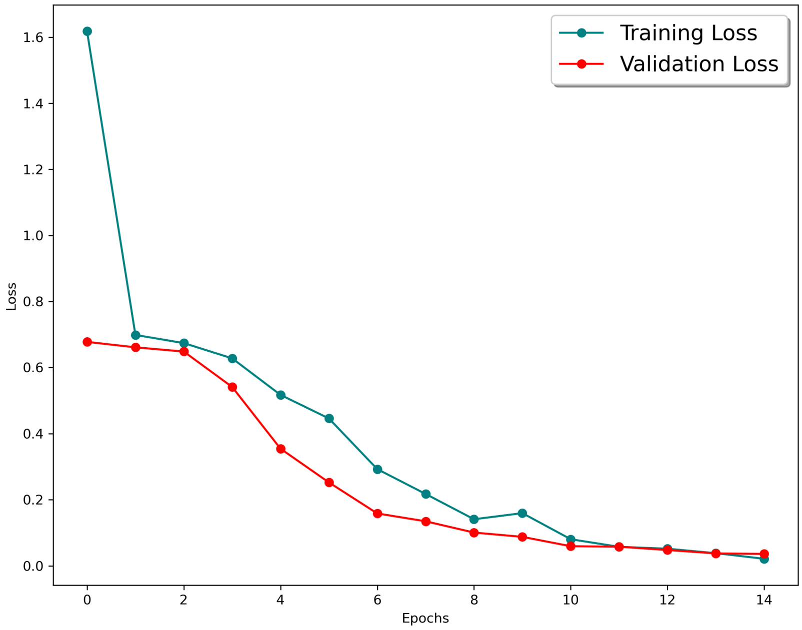

FIGURE 6. Training vs validation loss curve of CNN

The dataset is small and only contain two tribal dress. In future, we will add more tribal dresses, also different types of tribal dresses and different deep-learning algorithms will apply to find out the best classification matrix.

REFERENCES

[1] C. Eckert and M. Stacey, "Designing in the context of fashion–designing the fashion context," in *Designing in context: Proceedings of the 5th design thinking research symposium*, 2001, pp. 113–129.
[2] P. Dionisio, C. Leal, and L. Moutinho, "Fandom affiliation and tribal behaviour: a sports marketing application," *Qualitative Market Research: An International Journal*, vol. 11, no. 1, pp. 17–39, 2008.
[3] R. Setiawan, K. Rani, L. P. L. Cavaliere, N. T. Hiep, S. Halder, I. Raisal, R. Mishra, and S. S. Rajest, "References for shopping online versus in stores what do customers prefer and how do offline retailers cope with it?" Ph.D. dissertation, Petra Christian University, 2020.
[4] J. B. Griffin, "Eastern north american archaeology: A summary: Prehistoric cultures changed from small hunting bands to well-organized agricultural towns and tribes." *Science*, vol. 156, no. 3772, pp. 175–191, 1967.
[5] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, "Deepfashion: Powering robust clothes recognition and retrieval with rich annotations," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1096–1104.
[6] K. Lyu and H. Yan, "Identification method of dress pattern drawing based on machine vision algorithm," in *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*. IEEE, 2022, pp. 76–79.
[7] A. Rafaeli and M. G. Pratt, "Tailored meanings: On the meaning and impact of organizational dress," *Academy of Management Review*, vol. 18, no. 1, pp. 32–55, 1993.
[8] Y. Seo and K.-s. Shin, "Hierarchical convolutional neural networks for fashion image classification," *Expert systems with applications*, vol. 116, pp. 328–339, 2019.
[9] T. Mallavarapu, L. Cranfill, E. H. Kim, R. M. Parizi, J. Morris, and J. Son, "A federated approach for fine-grained classification of fashion apparel," *Machine Learning with Applications*, vol. 6, p. 100118, 2021.
[10] B. Willimon, I. Walker, and S. Birchfield, "A new approach to clothing classification using mid-level layers," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 4271–4278.

[11] J. Cychnerski, A. Brzeski, A. Boguszewski, M. Marmolowski, and M. Trojanowicz, "Clothes detection and classification using convolutional neural networks," in *2017 22nd IEEE international conference on emerging technologies and factory automation (ETFA)*. IEEE, 2017, pp. 1–8.

[12] P. B. Kowshik, A. V. Krishna, P. Reddy, and P. S. Sundar, "Classification of dress codes using convolution neural networks," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020, pp. 314–320.

[13] M. J. Hasan, M. S. Alom, U. F. Dina, and M. H. Moon, "Maize diseases image identification and classification by combining cnn with bi-directional long short-term memory model," in *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2020, pp. 1804–1807.

[14] R. Chauhan, K. K. Ghanshala, and R. Joshi, "Convolutional neural network (cnn) for image detection and recognition," in *2018 first international conference on secure cyber computing and communication (ICSCCC)*. IEEE, 2018, pp. 278–282.

[15] T. Jiang, X.-j. Hu, X.-h. Yao, L.-p. Tu, J.-b. Huang, X.-x. Ma, J. Cui, Q.-f. Wu, and J.-t. Xu, "Tongue image quality assessment based on a deep convolutional neural network," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–14, 2021.

[16] X. Zhen, J. Chen, Z. Zhong, B. Hrycushko, L. Zhou, S. Jiang, K. Albuquerque, and X. Gu, "Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study," *Physics in Medicine & Biology*, vol. 62, no. 21, p. 8246, 2017.

[17] M. A. Marjan, M. Hasan, M. Z. Islam, M. P. Uddin, and M. I. Afjal, "Masked face recognition system using extended vgg-19," in *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, 2022, pp. 1–4.

[18] S. P. Deore and A. Pravin, "Devanagari handwritten character recognition using fine-tuned deep convolutional neural network on trivial dataset," *Sādhanā*, vol. 45, pp. 1–13, 2020.

[19] M. Hasan, M. M. Islam, S. W. Sajid, and M. M. Hassan, "The impact of data balancing on the classifier's performance in predicting cesarean childbirth," in *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, 2022, pp. 1–4.

[20] O. Vesterinen, A. Toom, and S. Patrikainen, "The stimulated recall method and icts in research on the reasoning of teachers," *International Journal of Research & Method in Education*, vol. 33, no. 2, pp. 183–197, 2010.