

An Intelligent Fashion Object Classification Using CNN

Debabrata Swain^{1,*}, Kaxit Pandya¹, Jay Sanghvi¹ and Yugandhar Manchala²

¹Dept. of CSE, Pandit Deendayal Energy University, Gandhinagar, India

²Vardhaman College of Engineering, Hyderabad, India

Abstract

Every year the count of visually impaired people is increasing drastically around the world. At present time, approximately 2.2 billion people are suffering from visual impairment. One of the major areas where our model will affect public life is the area of house assistance for specially-abled persons. Because of visual improvement, these people face lots of issues. Hence for this group of people, there is a high need for an assistance system in terms of object recognition. For specially-abled people sometimes it becomes really difficult to identify clothing-related items from one another because of high similarity. For better object classification we use a model which includes computer vision and CNN. Computer vision is the area of AI that helps to identify visual objects. Here a CNN-based model is used for better classification of clothing and fashion items. Another model known as Lenet is used which has a stronger architectural structure. Lenet is a multi-layer convolution neural network that is mainly used for image classification tasks. For model building and validation MNIST fashion dataset is used.

Keywords: CNN, Lenet, Fashion items, Adam, ReLu, Fashion MNIST

Received on 09 August 2023, accepted on 31 October 2023, published on 06 November 2023

Copyright © 2023 D. Swain *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetinis.v10i4.4315

*Corresponding author. Email: debabrata.swain7@yahoo.com

1. Introduction

Fashion is among one of the world's most rapidly growing and ever-changing industries. With new trends and designs appearing every season, the fashion industry is always evolving. With the advent of technology and the widespread use of the internet, fashion has become more accessible to people worldwide. Fashion object detection is a rapidly growing field of computer vision that has garnered significant interest in the past few years due to its potential uses in various industries, like fashion retail, e-commerce, and social media. The ability to detect and classify different fashion objects such as clothing, accessories, and footwear in images has numerous practical benefits, including improved product search and recommendation systems, enhanced inventory management, and more effective marketing campaigns. In the fashion industry, object classification can aid in automating product categorization, improving online shopping experiences, and enabling more accurate visual search capabilities. This can help identify the type of

clothing, material, and color of an item in an image, retailers can better organize their inventory and offer personalized recommendations to customers. Additionally, social media platforms such as Instagram and Pinterest can use fashion object detection to enable more accurate tagging of products in images, making it easier for users to purchase items they see online. In past years, deep learning has come up as an important tool used for image recognition and its classification as well, allowing machines to recognize patterns in complex datasets with remarkable accuracy [1]. Deep learning models have been used in a wide range of fields, including robotics and autonomous systems as well as computer vision and natural language processing [2]. The main goal of this research is to create a deep learning model that can correctly identify a variety of clothing articles. This kind of application can be helpful for visually impaired individuals to recognize different fashion-related items. Convolutional Neural Networks (CNNs) [3] are among the type of deep learning model that is frequently used in image recognition tasks. These models consist of multiple layers of interconnected

neurons that learn to identify different features in images, such as edges, shapes, and textures. In this work, a fashion object recognition system has been developed using CNN.

2. Literature Review

Owais et. al. [4] focused on one of the hyperparameters of CNN which is the filter size. They recorded the Accuracy of 3x3, 5x5, and 7x7 filter sizes. Aside from the filter sizes everything is kept unchanged, with the model architecture remaining unchanged. The purpose of this experiment was to clarify how filter sizes affect image classification. 13 layered CNN was used here and they revealed that a 3x3 filter size was the best fit for the model. According to their experimental findings, accuracy, and filter size are inversely related. They have achieved an Accuracy of 92.68%. Though being less dense our proposed model gives better accuracy. Their proposed model uses Max Pooling while ours uses Average Pooling. Greeshma et. al. [5] performed several techniques to examine the effects of various regularisation and Hyper-Parameter Optimisation (HPO) strategies using deep neural networks. Two alternative neural network architectures having 2 ConvNets and 4 ConvNets are displayed in this work. Data Augmentation is also used along with deep ConvNet. The pooling function is also used here to resize the dimensions of the input image and also to prevent the model from being overfitted. This paper uses Adadelata as an optimizer along with Adam. Adam is more powerful than Adadelata which is one of the reasons our proposed system uses Adam as an optimizer. They have concluded that among all the models used, they have obtained the best accuracy in combination with CNN4, HPO, and Reg models. It achieves an Accuracy of 93% by Using classical CNN which is a bit less than ours as well. Tang et. al. [6] have studied in depth the roots of deep residual networks and worked in the direction of their optimization. It also includes the use of Residual Networks (RNN), Wide Residual Networks (WRN), and Pyramidal Residual networks (PyramidNet). The results gained here demonstrate that the model's performance can be enhanced by widening the network partly. This model takes more time and requires more computation power than our proposed system. They have found a combination of models, TTA technologies, Ranger, Auto-Augment, and CutMix, and is based on PyramidNet; which gives an accuracy was 96.21%. Kadam et. al. [7] explores the development of convolutional neural networks (CNNs) for image classification, focusing on their application to fashion classification. The review highlights five different architectures with varying convolutional layers, filter size and fully connected layers. This analysis is crucial to understanding the effectiveness of the proposed state-of-the-art model for fashion image classification, which achieves 92.76% accuracy with batch normalization using CNN. Greeshma et. al. [8] have

utilized one of the most straightforward and efficient single feature descriptors HOG. The HOG-based feature extraction method is used here for recognizing fashion products. One of the popular machine learning classifier techniques, multiclass SVM, is utilized to train the images. As a robust gradient-based feature descriptor that excels at data discrimination and is substantially superior to other feature sets, this combination is used. Feature Extraction is also used here. This work is performed in MATLAB and it has achieved an accuracy of 86.53%. Vijayaraj et. al. [9] have proposed a model with different activation functions, optimizers, learning rates, dropout rates, and batch sizes. Used deep neural network for Fashion based Image Classification. For extraction of the classification pattern different layers such as max, average and, sum pooling with fully connected layers are applied. The images are tested with ANN as well as the CNN. The highest Accuracy of 90%. Yusi Tang et. al. [10] have studied deep residual networks in detail and implemented them. Hyper-parameters' considerable effects on neural network models prompted research into the best way to build model structure and streamline the training process. Next, the training models' performances under various conditions were compared, and then the best training model was filtered which achieved an accuracy of 96.21%. Kexin Zhang et. al. [11] aims to develop and optimize a Recurrent Neural Network (RNN). To reduce the vanishing gradient problem and overcome RNN, LSTM's are used here. In addition to testing the Heuristic Pattern Reduction approach and the Network Pruning method, fine-tuning and cross-validation are also applied to optimize the model. Pytorch was used and they secured an accuracy of 89%. Bhatnagar et. al. [12] have used CNN as their deep learning model. For simplicity and learning process acceleration, they have developed three distinct convolutional neural network designs and utilized batch normalization and residual skip connections. By using a two-layer CNN along with Batch Normalization and Skip Connections They have attained an accuracy of 92.54%. Olivia Nocentini et. al. [13] have increased state-of-the-art accuracy by multiple convolutional neural networks with 15 convolutional layers (MCNN15). MCNN is a combination of Multiple Convolutional Neural Networks. Searching hyper-optimization and data augmentation techniques are also applied to improve the generalization of the models. It achieved a classification accuracy of 94.04%.

3. Proposed Work

3.1. Dataset description and pre-processing

Classical CNN and Lenet are the two models proposed here for the accurate identification of different fashion

items. MNIST fashion dataset is utilized for the training and validation of the model [14]. The size of the images is 28x28 pixels. Out of the total of 60,000 images, 5000 were considered for testing and 55,000 were used for training. The training samples are subdivided into training and validation. For validation, 5000 samples were taken and the rest were utilized for training. Each image pixel sp_i is further divided by 255 to normalize them between 0 and 1 as shown in equation 1. This normalization helps

the Neural Network to understand the features in a better manner. The dataset consists of fashion items which are classified into 9 categories.

$$P_i = \frac{sp_i}{255} \tag{1}$$

Where, $0 < sp_i < 1$.

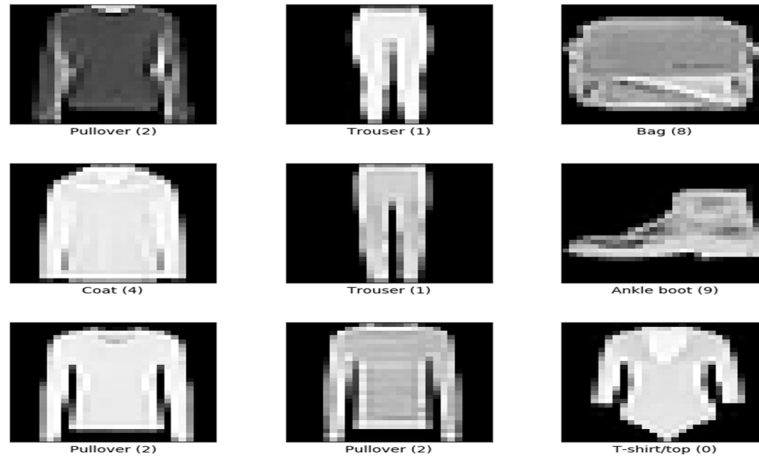


Figure 1: Sample Dataset Image

3.2. Proposed architecture

In this proposed work two models are used namely Classical CNN and Lenet as shown in figure 2 and 3. Classical CNN consists of different types of layers. Initially, a block of layers composed of the Convolution Layer, Dropout Layer, and Average Pooling layer is used twice. After that, another block comprised of Convolution, Dropout followed by a flatten, and finally

two dense layers are used. Except for the output layer, where the Softmax activation function is utilized, each convolution layer has used the relu activation function [15]. The second model used here is Lenet-5 which is an advanced version of CNN. It is the earliest pre-trained model. It has three different sets of convolutional layers, followed by average pooling. Two fully connected layers follow them. In the end, a Softmax classifier is used which classifies the images into respective classes [16]. The optimizer and loss function used in both models is Adam and sparse categorical cross-entropy respectively.

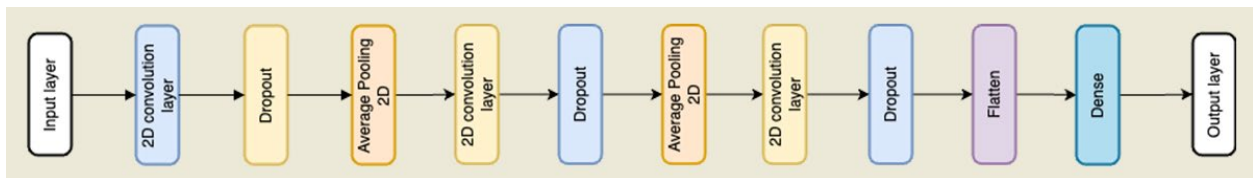


Fig 2: Classical CNN Architecture



Fig 3: LeNet Architecture

3.3. Proposed architecture

Rectified linear unit (ReLU)

ReLU ($R(x)$) has become one of the desired selections for a hidden layer's activation function because it is usually easier to teach and offers greater accuracy than others [17]. Additionally, it resolves the vanishing gradients, and its behavior is nearly linear as per equation 2.

$$f(x) = \max(0, x) \quad (2)$$

Softmax

It is also known as an exponential function, it is typically an output layer activation function [18, 19]. The function creates a probability distribution from the k input values in such a way that the sum of all the resulting output values is 1 as shown in equation 3.

$$E(z_i) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (3)$$

Where

e^{z_i} = Individual class output

k = Number of classes

3.4. Optimizer

Adam

Adam is a stochastic gradient descent algorithm used for training deep learning models. The learning rate used here was 0.001. One of the main reasons for its popularity is it can handle sparse gradients on noisy problems easily [20, 21] which we can observe from equation 4.

$$w_{\text{new}} = w_{\text{old}} - \eta \cdot (\hat{m} / \sqrt{\hat{s}} + \epsilon) \quad (4)$$

Where

$w_{\text{new}}, w_{\text{old}}$ = new and old weights

η = Learning rate

\hat{m} = Momentum term

\hat{s} = Scaling term

ϵ = Smoothing term to avoid zero division error

3.5. Loss Function

Sparse categorical cross-entropy

A low level of categorical cross-entropy calculates the difference between labels and forecasts [22]. It is mainly

utilized for multi-class classification models as a loss function where an integer value is assigned to the output label which can be seen in equation 5.

$$C = \sum_{i=1}^k y_i \ln(y_i) \quad (5)$$

Where

y_i = Output in i^{th} epoch

k = number of output classes

3.6. Dropout

Dropout is one of the most widely used Regularization Techniques [23]. The dropout function performs by arbitrarily dropping out unit activations for a single gradient step in a network [24]. Increasing dropout value results in a better model.

4. Results and Analysis

The accuracy of a classification model's predictions is gauged by its accuracy. In the figure 4 and 6, the accuracy of both the models during training and validation is depicted. The accuracy is the proportion between correctly predicted and total predicted images [25]. It is possible to achieve the traditional CNN model's training and validation accuracy at 91.64 and 93.76 respectively. The LENET model's training and validation accuracy is determined as 97.73 and 98.48 respectively as shown in Table 1.

The LENET model's prediction effectiveness increases with its increased precision. When compared to traditional CNN, LENET performs better overall because of its superior architectural design and wise choice of activation function, which includes ReLu, regularization techniques such as dropout, and the use of the most efficient optimizer which is Adam [26]. The accuracy and loss curve of both models is shown in figure 4, 5, 6 and 7. The Table 2 shows a comparative Analysis between the performance of the proposed models with the different literatures considered in this work. The Lenet model is computationally more efficient than the different models discussed in the table 2 because of its parameter efficiency and simplicity. The other reason for this significant performance of the model is due to the modern related parameter selections such as Activation function, Optimizer and Dropout. The Relu activation function and Dropout operation helped in avoiding the vanishing

gradient descent problem [27]. In this case, the Adam optimizer moved the gradient in the descent direction to reach the global minima. This reduction in the gradient value results in achieving better accuracy and reducing the loss. The pixel normalization operation also played a vital role during the process of model training. It helped the neural network to learn the features more easily from the images. This is also one of the key reasons to avoid overfitting of the proposed Neural Network structure.

Table1. Comparison between results of CNN and LeNet

Model Name	Validation Accuracy	Validation Loss	Training Accuracy	Training Loss
CNN	93.76	0.1714	91.64	0.2229
LeNet	98.48	0.0442	97.73	0.0615

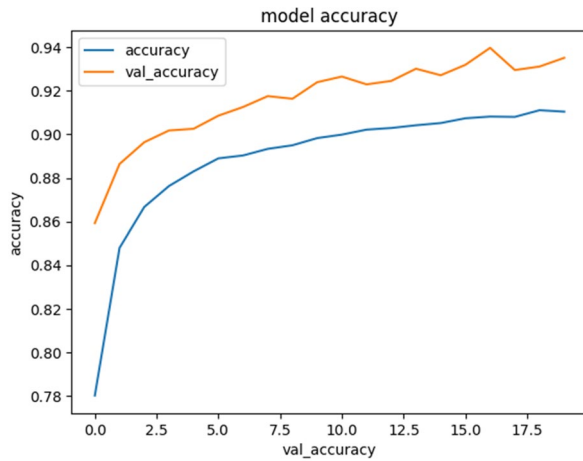


Fig 4: Accuracy curve for Classical

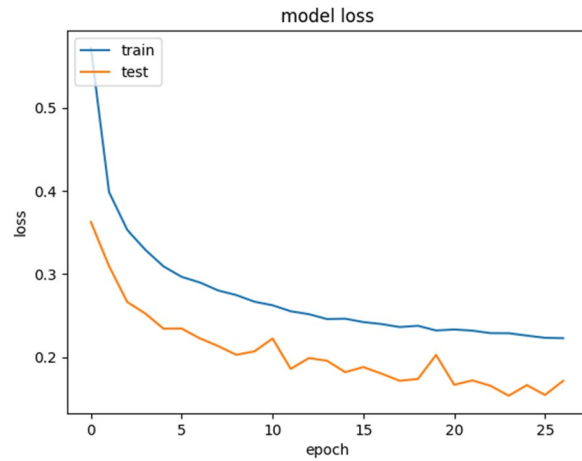


Fig 5: Loss curve for Classical CNN

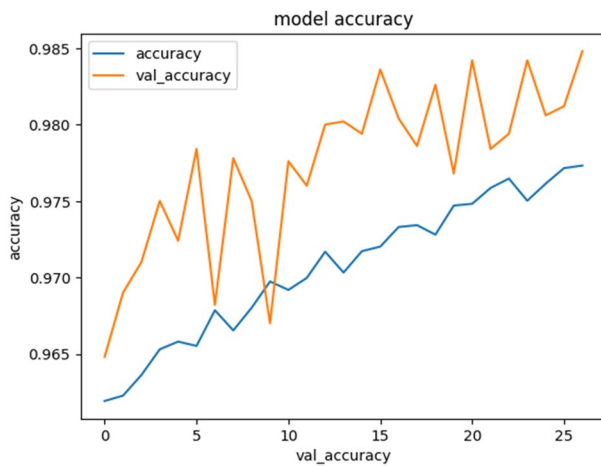


Fig 6: Accuracy curve for LeNet

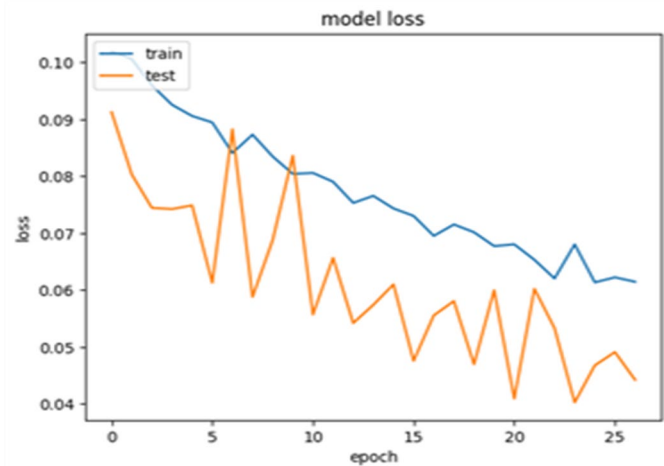


Fig 7: Loss curve for LeNet

Table. 2: Comparative Analysis Table

Reference number	Model	Accuracy
4	CNN with variable Filter	92.6%
5	CNN with Hyperparameter optimization	93%
6	PyramidNet	96%
7	CNN	92.5%
8	HOG based Feature Extraction	86.5%
9	Tunned CNN	90%
10	Residual network with Hyper Parameter Tunning	96.2%
11	Optimize RNN	89%
12	Two Layer CNN	92.5%
13	multiple convolutional neural networks	94%
	Proposed System (CNN)	93.7%
	Proposed System (Lenet)	98.4%

5. Conclusion

Our tests show that CNN and LeNet models are successful in classifying the Fashion MNIST dataset. The LeNet model achieved a higher accuracy than the CNN model, indicating that deeper and more complex architectures can yield better results for image classification tasks.

The results also suggest that the Fashion MNIST dataset is a challenging but reliable benchmark for evaluating the performance of image classification models. Future research could explore the use of other deep learning architectures and transfer learning techniques to further improve the classification accuracy of the dataset. A detailed investigation has been done on deep learning-based models for the classification of several fashion items. It can be used for robot assistance systems which could help visually impaired people as well as people of old age. The Proposed model has shown higher accuracy as compared to other models.

References

- [1] Christopher Geier, Training on test data: Removing near duplicates in Fashion-MNIST, DOI: <https://doi.org/10.48550/arXiv.1906.08255>
- [2] Han Xiao, Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms, DOI: <https://doi.org/10.48550/arXiv.1708.07747>
- [3] E. Xhaferri, E. Cina and L. Toti, Classification of Standard FASHION MNIST Dataset Using Deep Learning Based CNN Algorithms, 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 2022, pp. 494-498, doi: 10.1109/ISMSIT56059.2022.9932737.
- [4] Khanday, Owais & Dadvandipour, Samad & Lone, Mohd Aaqib. (2021). Effect of filter sizes on image classification in CNN: a case study on CFIR10 and Fashion-MNIST datasets. IAES International Journal of Artificial Intelligence (IJ-AI). 10. 872. 10.11591/ijai.v10.i4.pp872-878.
- [5] K V, Greeshma & K., Sreekumar. (2019). Hyperparameter Optimization and Regularization on Fashion-MNIST Classification. International Journal of Recent Technology and Engineering. 8. 3713. 10.35940/ijrte.B3092.078219.
- [6] Tang, Yusi & Cui, Hanguo & Liu, Shuyong. (2020). Optimal Design of Deep Residual Network Based on Image Classification of Fashion-MNIST Dataset. Journal of Physics: Conference Series. 1624. 052011. 10.1088/1742-6596/1624/5/052011.
- [7] Kadam, Shivam & Adamuthe, Amol & Patil, Ashwini. (2020). CNN Model for Image Classification on MNIST and Fashion-MNIST Dataset. Journal of scientific research. 64. 374-384. 10.37398/JSR.2020.640251.
- [8] K V, Greeshma & K., Sreekumar. (2019). Fashion-MNIST classification based on HOG feature descriptor using SVM. International Journal of Innovative Technology and Exploring Engineering. 8. 960-962.
- [9] Vijayaraj et al, Deep Learning Image Classification for Fashion Design, Wireless Communications and Mobile Computing, 2022, DOI: <https://doi.org/10.1155/2022/7549397>
- [10] Tang, Yusi & Cui, Hanguo & Liu, Shuyong. (2020). Optimal Design of Deep Residual Network Based on Image Classification of Fashion-MNIST Dataset. Journal of Physics: Conference Series. 1624. 052011. 10.1088/1742-6596/1624/5/052011.
- [11] Zhang, K. (2018), LSTM: An Image Classification Model Based on Fashion-MNIST Dataset, Doi: <https://api.semanticscholar.org/CorpusID:21102649>
- [12] Bhatnagar, Shobhit & Ghosal, Deepanway & Kolekar, Maheshkumar. (2017). Classification of fashion article images using convolutional neural networks. 1-6. 10.1109/ICIP.2017.8313740.
- [13] Nocentini O, Kim J, Bashir MZ, Cavallo F. Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset. Sensors. 2022; 22(23):9544. <https://doi.org/10.3390/s22239544>
- [14] https://www.tensorflow.org/datasets/catalog/fashion_mnist

- [15] Swain, Drdebabrata & Satapathy, Santosh & Acharya, Biswaranjan & Shukla, Madhu & Gerogiannis, Vassilis & Kanavos, Andreas & Giakovis, Dimitris. (2022). Deep Learning Models for Yoga Pose Monitoring. *Algorithms*. 15. 403. 10.3390/a15110403.
- [16] Mr. Debabrata, D. S. Kumar, and Dr Debabala, "Diagnosis of Coronary Artery Disease using 1-D Convolutional Neural Network," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2. Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP, pp. 2959–2966, Jul. 30, 2019. doi: 10.35940/ijrte.b2693.078219.
- [17] Debabrata Swain et al, Improved handwritten digit recognition using artificial neural networks, *International Journal of Computing Science and Mathematics*, 2023 Vol.17, No.4, DOI: 10.1504/IJCSM.2023.131625
- [18] Kumar, S., Neware, N., Jain, A., Swain, D., Singh, P. (2020). Automatic Helmet Detection in Real-Time and Surveillance Video., *Machine Learning and Information Processing. Advances in Intelligent Systems and Computing*, vol 1101. Springer, Singapore. https://doi.org/10.1007/978-981-15-1884-3_5
- [19] Chao Duana et al., Image Classification of Fashion-mnist Data Set Based on VGG Network, 2019 2nd International Conference on Information Science and Electronic Technology (ISET 2019), DOI: 10.23977/iset.2019.004
- [20] M. Y. W. Teow, Experimenting Deep Convolutional Visual Feature Learning using Compositional Subspace Representation and Fashion-MNIST, 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 2020, pp. 1-6, doi: 10.1109/IICAIET49801.2020.9257819.
- [21] D. Swain et al, Diabetic retinopathy using image processing and deep learning, *International Journal of Computing Science and Mathematics*, Vol. 14, No. 4, pp 397-409, 2022, doi.org/10.1504/IJCSM.2021.120686.
- [22] D. Swain et al., "A Deep Learning Framework for the Classification of Brazilian Coins," in *IEEE Access*, vol. 11, pp. 109448-109461, 2023, doi: 10.1109/ACCESS.2023.3321428.
- [23] Swain, D., Parmar, B., Shah, H. et al. Enhanced handwritten digit recognition using optimally selected optimizer for an ANN. *Multimed Tools Appl* 82, 44021–44036 (2023). <https://doi.org/10.1007/s11042-023-15402-0>
- [24] J. Shen and M. O. Shafiq, "Deep Learning Convolutional Neural Networks with Dropout - A Parallel Approach," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 2018, pp. 572-577, doi: 10.1109/ICMLA.2018.00092.
- [25] Hossin, Mohammad & M.N, Sulaiman. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. 5. 01-11. 10.5121/ijdkp.2015.5201.
- [26] Lydia, Agnes & Chandrasekar, Sheela. (2022). A Comparative Study on Regularization Techniques in Convolutional Neural Networks. 784-793.
- [27] H. H. Tan and K. H. Lim, "Vanishing Gradient Mitigation with Deep Learning Neural Network Optimization," 2019 7th International Conference on Smart Computing & Communications (ICSCC), Sarawak, Malaysia, 2019, pp. 1-4, doi: 10.1109/ICSCC.2019.8843652.