Science World Journal Vol. 19(No 2) 2024 https://dx.doi.org/10.4314/swj.v19i2.27
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Full Length Research Article

CLOTHING IMAGE CLASSIFICATION MODEL USING REGULARIZED MULTIPLE CONVOLUTIONAL NEURAL NETWORKS (RMCNN)

*[1]Shuwa Ayongo, [1]Muhammad Abdullahi, [1]Mustapha Aminu Bagiwa, [2]Alice Oluwasikemi Matemilola

[1]Department of Computer Science, Ahmadu Bello University, Zaria
[2]Department of Computer Science, Federal College of Education, Zaria

*Corresponding Author Email Address: shuwajunior@gmail.com

ABSTRACT

In the past few decades, machines have gradually taken over the daily activities of human beings such as online shopping and clothes manipulation. It is essential to develop artificial intelligence techniques that can help people detect and classify clothing designs accordingly. Early efforts to solve the clothing image classification problem require carefully selecting and extracting certain features from clothing image datasets in such a way that the features of the datasets are highly represented. However, these methods are difficult in defining and capturing a wide range of image features. Research shows that Convolutional Neural Networks (CNN) models can solve image classification problems better than traditional machine learning (ML) methods. However, they are faced with problems such as over-fitting, Hyper-parameter tuning, Noisy data, and insufficient training data. This work addresses the problem of overfitting which reduces the classification/generalization performance of clothing image classification models. We proposed four (4) CNN models in which a Regularization method called Dropout is added to each layer to handle the over-fitting problem. The model with the best result out of the four is adopted as the proposed model. The results show about 1.77% improvement in accuracy as compared to the results recorded by other models that were trained using the same dataset and the state-of-the-art architectural designs.

Keywords: Regularization, Neural Network, Classification, Artificial Intelligence, Computer vision, Overfitting, Dropout

INTRODUCTION

In the past few decades, machines are gradually taking over the daily activities of human beings such as online shopping and clothes manipulation. Online shopping and Clothes manipulation need certain features such as colour, design, and shape of the clothes for the machines to be able to identify and group them accordingly. It is essential to develop artificial intelligence techniques that can detect and classify clothing designs appropriately to enable machines perform the task of online purchasing as well as helping them decide the type of clothes for humans effectively. This can help users understand the products better and attract more customers from different locations thereby boosting sales. A greater understanding of the customers' tastes, cultures, and socioeconomic position may also be aided by this kind of information(Henrique et al., 2021).

Classification of clothing fashion designs belongs to the broader group in computer vision called image classification. The task of classifying objects into various categories can be considered as an easy one for humans but complex one for machines. It is one of the most important but challenging tasks in various application domains such as medical imaging, object identification, traffic control systems, brake light detection, machine vision, video surveillance, vehicle navigation, industrial visual inspection, remote sensing, robot navigation, Fashion, and design (Nocentini et al., 2022).

Image classification refers to the process of grouping images into any of the different predefined categories. Early efforts to solve the image classification problem require carefully selecting and extracting certain features from image datasets in such a way that the features of the datasets are highly represented. The features are normally created to capture relevant information in images like colour, text, shapes, or a combination of both. The feature extraction is achieved using Machine learning classifiers like Support Vector Machine (SVM), Naïve Bayes, Decision Tress (DT), K-Nearest Neighbours (KNN), Random Forest (RF), and Multilayer Perceptron (MLP). However, these methods were difficult in defining and capturing a wide range of information from images. This gives birth to the Convolutional Neural Network (CNN) which provides better ways of learning these features(Nocentini et al., 2022).

According to Nocentini et al., (2022), a Convolutional neural network is a special type of multi-layer neural network inspired by the mechanism of the optical system of living creatures. Convolutional neural network in image classification aims to learn the output features directly from data and to use it to classify images according to their predefined classes. The basics of CNNs consist of three stages, namely; the convolutional layer, non-linearity (Activation layer), and the Pooling layer.

In as much as CNN performance in computer vision is rated above the traditional machine learning algorithms, it is faced with some challenges such as overfitting and its inability to generalize which need to be addressed(Ying, 2019).

According to Ying, (2019), overfitting is an undesirable behaviour that occurs when a machine learning model gives accurate predictions for training data but fails to do the same to new or unseen data. It is caused by insufficient data, Noisy data, complex architectural designs, high image resolution, etc. Overfitting causes deep learning models to perform poorly on data that it has not seen before. It can be controlled though regularization, early stopping, and ensemble methods. Precisely, this work addresses the problem of overfitting in clothing image classification through a Regularization method called Dropout.

LITERATURE REVIEW

Fashion image classification is an area of computer vision and deep learning that tries to categorize and label fashion related

Science World Journal Vol. 19(No 2) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i2.27

images based on their visual content. It has become an essential technique for online retailers, fashion designers, and people that have passion about fashion (Nocentini et al., 2022).

Like other deep learning tasks, image classification is faced with several challenges. The main challenges of fashion image classification come from the wide diversity of fashion items, variations in garment positions, occlusions, variable lighting situations, and complicated backgrounds(Vijayaraj et al., 2022).

It is interesting to note that there has been a lot of work done in the past to classify fashion images using different machine learning and deep learning techniques as well as Fashion-MNIST image datasets (Eshwar et al., 2016). According to Ying, (2019), some of the techniques include Regularization, early stopping, and Ensemble methods. Nocentini, et al., (2022), itemized some of the datasets used for fashion image classification as AG dataset, DeepFashion-C dataset, IndoFashion dataset, Fashion-MNIST dataset, Fashion-Product. Some of the works done in the area of fashion image classification are presented in this work based on the dataset used as well as the machine/deep learning technique used.

Donati et al., (2019) conducted a real-world study using AG dataset. The work aims to automatically recognize and classify Logos, stripes, colours, and other features of the clothes.

In their research work, Su et al., (2020) proposed a framework to retrieve fashion products using DeepFashion-C dataset. The work was inspired from the Biological human brain and was targeted at extracting landmark localization data as well as fine-tuning the way the information will appear. Similarly, Shajini & Ramanan, (2021) used the DeepFashion-C to improve the classification accuracy of clothing images. They integrated two attention pipelines, landmark-driven, and spatial-channel attention. Their model improves the classification accuracy by representing the multi-scale landmark contextual information and locating the most important features of the input image.

In later work, Shajini & Ramanan, (2022) performed a study that focused on enhancing the feature representation by simultaneous learning of labelled and unlabelled samples. The study, offered a semi-supervised multi-task learning strategy with the goal of achieving attribute prediction and apparel category classification. They use a teacher-student (T-S) pair model that uses weighted loss minimization while exchanging information between teacher and student in order to increase semi-supervised fashion clothing analysis.

Some authors centred their findings on the Fashion-MNIST which is one among the numerous datasets that has been used for classification of Fashion images. In most cases, the accuracy achieved in certain instances exceeds 90% (Nocentini et al., 2022). Other authors used this dataset to perform the classification of images and their various findings are recorded in literature.

Eshwar S G et al., (2016) presented a fresh approach to online clothing purchasing. Their method of categorizing images using convolutional neural networks focuses mostly on Categorization of the type of clothing into many classes; and retrieval of similar clothing using the query image. The work was implemented using CNN to retrain the final layer of a GoogleNet to classify the clothing images.

In order to facilitate and accelerate the learning process, Bhatnagar et al., (2017), proposed three different convolutional neural network architectures and used batch normalization and residual skip connections. They trained the convolutional neural network-based on deep learning architectures to classify images in the Fashion-MNIST dataset. Their result showed a 2% increment in accuracy when compared with the state of the art recorded in literature.

With the aim of improving the classification accuracy of the classification models, Henrique et al., (2021) used the Fashion-MNIST dataset and proposed four distinct convolutional neural network models. The result of their model shows significant increase in the classification accuracy as compared to the results in the existing state of the art literature.

Seo & Shin, (2019), suggested classifying clothing materials using hierarchical convolutional neural networks (H-CNN). The study is significant since it is the first to apply hierarchical classification of clothing using CNN, and it makes a contribution because the suggested model is a knowledge embedded classifier that produces hierarchical data. They used the Fashion-MNIST dataset to develop H-CNN using VGGNet. The accuracy and loss are both reduced when employing the H-CNN model compared to the base model without a hierarchical structure, according to the results, H-CNN performs better when classifying clothing items.

Duan et al., (2022) was able to swap out the larger convolution cores (11x11, 7x7, and 5x5) in AlexNet with a series of sequential 3x3 convolution cores. Small cumulative convolution kernels outperform large convolution kernels for a particular receptive field. Increasing network depth with multilayer nonlinearity layer ensures learning of increasingly complicated patterns at a low cost. In addition, a batch normalization layer is added after each pooling layer to standardize input data and create a more uniform distribution of features across all features, making it simpler to train efficient models. Finally, 91.5% of the Fashion-MNIST Data Set classifications made in this article were accurate.

Lei et al., (2020), suggested a unique shallow convolutional neural network (SCNNB) that makes use of batch normalization approaches to speed up training convergence and enhance accuracy in order to get over challenges of consuming computing resources and time wasting. Their work achieved a general classification accuracy of 93.69%.

With normalization, some tuning, and reduction of overfitting, Saiharsha et al., (2020) reported that they achieved an accuracy of 91.78% with a VGG-like architecture; they also tested a CNN with the Fashion-MNIST dataset and achieved nearly the same test accuracy (90.77%); the authors noted that VGG produces better results but at the cost of taking a long time to train and being more computationally intensive.

Greeshma & Sreekumar, (2019) demonstrated the classification of the Fashion-MNIST dataset using the multiclass Support Vector Machine (SVM) and the Histogram of Oriented Gradient

Science World Journal Vol. 19(No 2) 2024
https://dx.doi.org/10.4314/swj.v19i2.27
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

(HOG) feature descriptor. They examined how one popular feature descriptor affects classification tasks for fashion products. According to the authors, they used HOG which is one of the most straightforward and useful single feature descriptors to train the images using one of the top multiclass machine learning classifier methods called SVM.

On the Fashion-MNIST and CIFAR-10 datasets, Hoang, (2019) compared the performance of various models including SVM, K-Nearest Neighbour, Random Forest, Decision Tree, and CNN. The research also looked at several feature extraction methods to enhance the model's functionality. The results show that the strategy of employing an auto-encoder was superior to the Principal Component Analysis (PCA) for improving the model's performance; specifically, the results showed that combining an auto-encoder with SVM outperformed a CNN model.

Using the Fashion-MNIST dataset, hyper-parameter optimization, and regularization methods such as image augmentation and dropout are utilized to increase the accuracy of networks in classification tasks. This method was utilized by Greeshma & Sreekumar, (2019) using four-layer ConvNets to achieve an accuracy of 93.99%.

From the preceding paragraphs, there are many approaches from the literature that were utilized to train the Fashion-MNIST dataset and showed excellent results when tested, but there is still potential for improvement in terms of the accuracy of this kind of dataset.

In order to increase the image classification accuracy, Nocentini et al., (2022) used Hyper-parameter optimization and data augmentation techniques to improve the classification accuracy of Fashion images. The work achieved overall classification accuracy of 94.04% on the Fashion-MNIST dataset.

The review of the related literature shows that there are existing gaps that needs to be addressed. Specifically, in terms of classification accuracy. It is observed that in the work of Nocentini et al., (2022), the model struggles to clearly classify some dataset items, which according to Thanapol et al., (2020), is due to the fact that both Hyper-parameter optimization and data augmentation methods suffers from the problem of overfitting which makes networks to perform poorly in terms of classification accuracy and generalization. This work therefore proposed a model that would control overfitting in clothing image classification and improve the general clothing classification accuracy with improved generalization performance.

MATERIALS AND METHODS

Dataset

This paper uses Fashion-MNIST dataset. It is a dataset designed to represent modern computer vision tasks, and it is publicly available to the global research community on https://www.kaggle.com/datasets/zalando-research/fashionmnist. The Fashion-MNIST dataset was created by Zalando with the aim of replacing the original MNIST dataset of handwritten digits, which is powered by renewable energy. It comprises of seventy thousand (70,000) grey scale images, comprising of sixty thousand (60,000) training images, and ten thousand (10,0000) validation images. The size of each of the images is 28 x 28 pixels. The images are grouped into ten (10) classes, namely: T-shirt/Top, Trousers, Pullover, Dress, Coat, Sandal, Shirt, Sneakers, Bag, and Ankle boot. The ten categories of the Fashion MNIST dataset is shown in Figure 1
Figure 1: Sample images of the Fashion MNIST dataset

System Architecture

The architecture of the proposed system is based on the Multiple Convolutional Neural Network (MCNN) model proposed by Nocentini et al., (2022). It is aimed at solving the problem of overfitting associated with most of the CNN models leading to poor generalization performance.

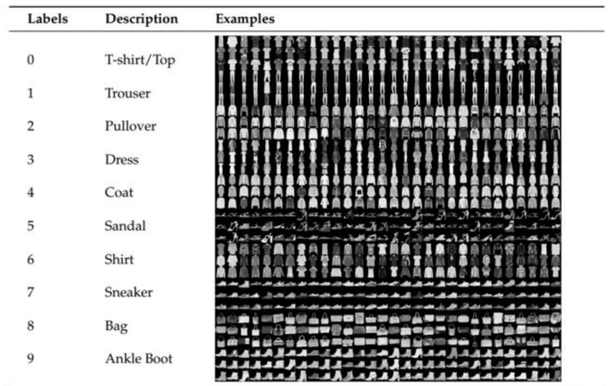

Figure 1: Sample images of the Fashion MNIST dataset

The proposed model is called Regularized Multiple Convolutional Neural Network Model (RMCNN).
According to Ying, (2019), overfitting can be controlled through simpler architectural designs and a regularization method called Dropout. Based on this fact, four different architectural CNN models namely; RMCNN1, RMCNN2, RMCNN3, and RMCNN4 were proposed. A dropout layer was added to each of the network groups to control overfitting. The four proposed models were tested individually using the Fashion-MNIST dataset. The result of the best model out of the four models is selected and compared with other architectural designs as well as other models built using Fashion MNIST dataset.

The Regularized Multiple Convolutional Neural Network (RMCNN) Model Architecture

The design of our four models (RMCNN1, RMCNN2, RMCNN 3, and RMCNN4) is based on those of Nocentini et al., (2022) which is our base model.

2.0 **RMCNN1 and RMCNN2:** These models comprise of the four (4) groups of convolutional, MaxPooling, and Dropout layers. The architectural trend in the layers is based on Conv-MaxPooling-Dropout-Conv-MaxPooling-Dropout. The Dropout layer is added to handle overfitting which is always the problem that affects the performance of a neural network. A Fully Connected (FC) layer and a Softmax layer are added after the four groups of layers. The only difference between the RMCNN1 and RMCNN2 is the trends in the number of channels. The RMCNN1 uses 32-32-64-64, while RMCNN2 uses 32-64-128-256 and the number of convolutional groups. While the RMCNN1 consists of five (5) convolutional groups, the RMCNN2 is made up of six (6) convolutional groups.

RMCNN3 and RMCNN4: These comprise of the four (4) groups of convolutional, MaxPooling, and Dropout layers. The architectural trend in the layers is based on Conv-Conv-MaxPooling-Dropout-Conv-Conv-MaxPooling-Dropout. A Fully Connected (FC) layer and a Softmax layer of ten (10) nodes are added after the four groups of layers. The only difference between the RMCNN3 and RMCNN4 is the trends in the number of channels. The RMCNN3 uses 8-16-32-64-128-256-512-1024, while RMCNN4 uses 32-32-64-64-128-128-256-256. The full architectural design of the proposed model is shown in Figure 2

Regularization

Regularization is a technique used in deep learning to prevent the problem of overfitting by adding constraints to the model's optimization process, thereby making the model to generalize better and avoid fitting noise in the training data.

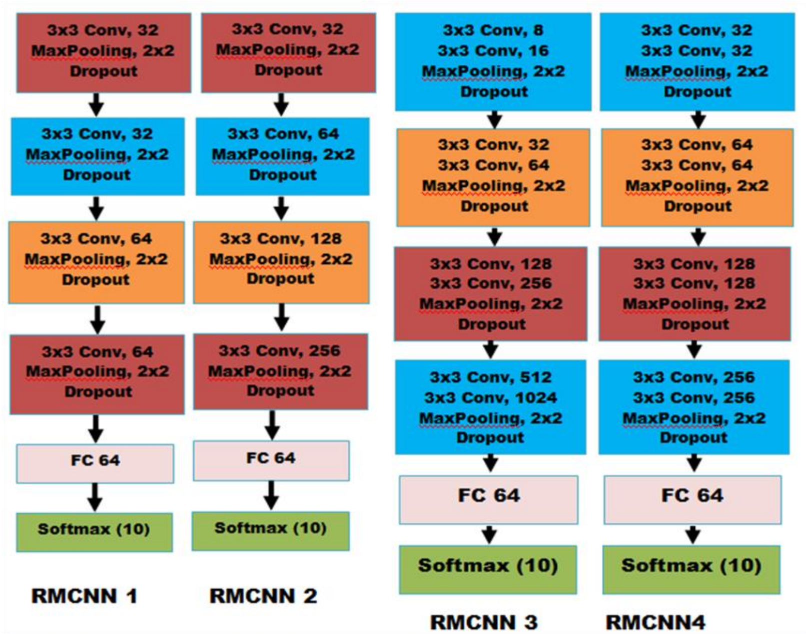

Figure 2: Architecture of the proposed model

Some common regularization methods include L1 Regularization, L2 Regularization, and Dropout (Ying, 2019).
L1Regularization (Lasso Regularization) modifies the loss function by adding a penalty term proportionate to the absolute values of the model's weights. This promotes the model to have sparse weight values, causing some of them to be absolutely zero. It is beneficial for feature selection since it directs the model's attention to a subset of the most significant features. L2 Regularization also known as Ridge Regularization on the other hand adds a penalty term proportionate to the squared values of the model's weights. As a result, heavy weights are discouraged and the model's weights are dispersed more evenly. In order to keep the model from becoming unduly sensitive to minute changes in the input data, L2 regularization is useful (Srivastava et al., 2014).
According to Ranjan, (2020), as Deep Learning architectures are becoming deeper and more comprehensive with even better classification accuracies, it was discovered that regularizing the network with L1 and L2 regularization methods was not able to solve the co-adaptation problem.
Co-adaptation occurs when regularization techniques fail to prevent the correlation between the weights of different neurons in a neural network. Co-adaptation can lead to overfitting and poor generalization performance of the model. L1 and L2 regularization, which aims to reduce the magnitude of the weights, does not prevent co-adaptation. This problem made Deep Learning not to be famous. The concept of Dropout which was introduced around 2012 transformed deep Learning due to its ability to control overfitting(Ranjan, 2020).
Dropout is a regularization technique that is specifically used for neural networks. It controls overfitting by randomly dropping out the neurons (nodes) and their connections out of the network during training thereby setting their activations to zero. This pushes the network to learn more robust properties by preventing neurons from co-adapting(Srivastava et al., 2014).
According to Hinton et al., (2012), overfitting is greatly reduced by randomly omitting half of the feature detectors on each training case. This avoids complicated co-adaptations where a feature detector is useful only when combined with multiple other particular feature detectors. Rather, given the combinatorial wide range of internal settings in which it must function, each neuron learns to recognize a trait that is typically helpful for delivering the right answer. Similarly, Baldi & Sadowski (2014), stated that in its most basic version, each feature detector unit is simply randomly eliminated with probability q = 1, p = 0.5 on each presentation of each training sample. Back-propagation training is used to train the remaining weights. For each example and training epoch, the process is repeated while sharing the weights. Predictions are generated by halving all the weights following the training session as shown in Figure 3

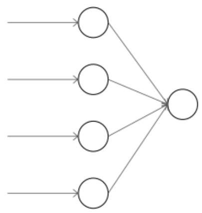

Figure 3(a): A Full Network

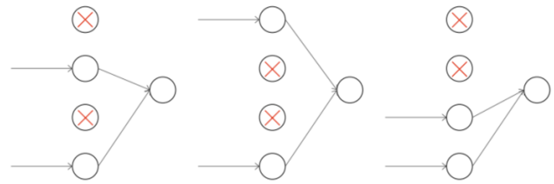

Figure 3(b): A Regularized Network

The probability of setting activation to zero is determined by the dropout rate, which is typically set between 0.2 and 0.5 according to equation 1

q= 1 – p .... (1)

Where p = probability

Considering a single layer network, the output of the layer is a linear weighted sum of inputs.

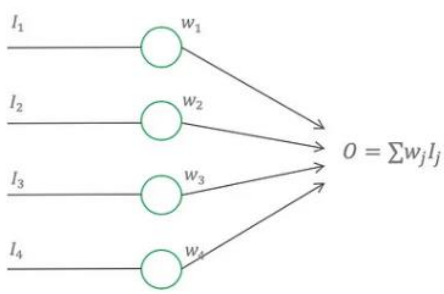

Figure 4: A Single layer linear unit out of network.

By estimating the model, the loss function can be minimized according to equation 2

$$E_n = \frac{1}{2}(t - \sum_{1}^{n} w_i I_i)^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

(2)

Where w = weight of Input, I = Input vector

When the dropout rate is added to the normal network unit, the dropout equation will be obtained.

$$E_n = \frac{1}{2}(t - \sum_{1}^{n} \delta_i w_i I_i)^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

(3)

For a regularized network, the dropout equation is given as:

$$E_n = \frac{1}{2}(t - \sum_{i=1}^{n} \rho_i w_i I_i)^2 + \sum_{i=1}^{n} \rho_i (1 - \rho_i) w_i^2 \;\; I_i^2 \dots\dots\dots\dots\dots\dots\dots$$

(4)

Implementation Procedure

After creating the basic models, Tensorflow and Keras Libraries of the Python Programming Language were used together with Fashion-MNIST dataset to train the models. The models were compiled. The loss was set to sparse categorical cross-entropy and metrics as sparse categorical accuracy. During the models' training, the epochs were set to 20, and steps per epoch as 100. Adam Optimizer was used for all the four models at a learning rate of 0.001. The models' validation split was set at 33% to get better test accuracy and have a minimum loss.

.

Evaluating the Proposed Model

The performance of the proposed model was tested using Fashion MNIST dataset. Its performance metrics were based on Accuracy, Precision, Recall, and F1-Score. A confusion matrix was used to analyze the above-mentioned metrics.

The result obtained would be compared to that of our base model and state-of-the-art performance recorded in the related literature.

Accuracy

According to Amin K.; et al (2019), accuracy is the degree to which a computer model or algorithm correctly assigns images to the appropriate categories. Generally, a tagged collection of images (dataset) with their corresponding actual labels are required to calculate the accuracy of image classification. Calculating the proportion of images that were properly classified out of all the images in the testing subset allows one to assess the accuracy of the model's predictions. The accuracy is typically expressed as a percentage.

Mathematically, it can be expressed as

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (5)$$

Where

TP = True Positive, TN = True Negative

FP = False Positive, FN = False Negative

Precision

Precision is a commonly used metric in image classification models to evaluate the accuracy of positive predictions made by the model. It measures the proportion of true positive (TP) predictions (correctly identified positive cases) among all the instances the model predicted as positive (Brownlee, J. 2020)

Mathematically, it can be expressed as:

$$\text{Precision} = \frac{TP}{(TP + FP)} \qquad (6)$$

Where

TP = True Positive, FP = False Positives

A high precision value indicates that each time the model predicts an image as positive, it is usually correct, which is important for tasks where false positives are costly or undesirable.

Recall

Recall is also known as sensitivity or True Positive rate. It is another important metric for evaluating the performance of an image classification model. Recall measures the proportion of true positive predictions (correctly identified positive cases) out of all the actual positive instances in the dataset.

The formula for calculating recall is as follows:

$$\text{Recall} = \frac{TP}{(TP + FN)} \qquad (7)$$

Where:

Science World Journal Vol. 19(No 2) 2024 https://dx.doi.org/10.4314/swj.v19i2.27
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

TP = True Positives (TP), FN = False Negatives (FN)

F1-Score

The F1-score is a commonly used metric in image classification models, which combines both precision and recall into a single value, providing a balanced evaluation of a model's performance. It is especially useful when you want to strike a balance between minimizing false positives and false negatives. The F1-score is the harmonic mean of precision and recall.

The formula for calculating the F1-score is as follows:

$$F1 - Score = 2\ X \frac{(Precision\ x\ Recall)}{(Precision + Recall)} \quad (8)$$

The F1-score is particularly valuable when the costs associated with false positives and false negatives are not equal, and you need to find a trade-off that works best for your specific application.

RESULTS AND DISCUSSION

Performance of the Proposed Models

The performances of the four proposed models were considered individually using the same dataset, learning rate, and number of epochs, and the results of the various models were recorded for comparison based on Accuracy, F1-score, Recall, and Precision.

Performance of the RMCNN1 Model

The RMCNN1 which is the first out of the four proposed CNN models was tested using Fashion MNIST dataset. The result of its classification performance for each of the ten items in the dataset is visualized as shown in Figure 5

It produces the overall classification accuracy of 90.39%. The result obtained from other evaluation metrics are as presented in

Figure 5: Visualization of RMCNN1

Table I
Table I: Performance of RMCNN1

	Precision	**Recall**	**F1-Score**
T-shirt/Top	88%	81%	85%
Trousers	99%	97%	98%
Pullover	85%	84%	85%
Dress	90%	90%	90%
Coat	82%	83%	82%
Sandals	99%	97%	98%
Shirt	69%	74%	71%
Sneakers	94%	96%	95%
Bag	97%	98%	97%
Ankle boot	96%	96%	96%

Performance of the RMCNN2 Model

The second proposed and implemented model is the RMCNN2. It was also tested using Fashion MNIST dataset. Its classification performance on the individual dataset items is visualized as shown in Figure 6

It produces the overall classification accuracy of 91.98%. Other evaluation measures of the RMCNN2 model are presented in Table II.

Figure 6: Visualization of RMCNN2

Table II: Performance of the RMCNN2 model

	Precision	Recall	F1-Score
T-shirt/Top	87%	85%	86%
Trousers	100%	99%	99%
Pullover	85%	90%	87%
Dress	94%	93%	94%
Coat	91%	83%	87%
Sandals	99%	98%	98%
Shirt	74%	79%	76%
Sneakers	94%	99%	96%
Bag	98%	99%	99%
Ankle boot	99%	95%	97%

Performance of the RMCNN3 Model

The RMCNN3 is the third model proposed. After execution, it recorded the overall classification accuracy of 95.81%. It is visualized as shown in Figure 7

Figure 7: Visualization of RMCNN3

The details of other performance measures are presented in Table III

Table III: The Performance of the RMCNN3 model

	Precision	Recall	F1-Score
T-shirt/Top	87%	85%	86%
Trousers	100%	99%	99%
Pullover	85%	90%	87%
Dress	94%	93%	94%
Coat	91%	83%	87%
Sandals	99%	98%	98%
Shirt	74%	79%	76%
Sneakers	94%	99%	96%
Bag	98%	99%	99%
Ankle boot	99%	95%	97%

Science World Journal Vol. 19(No 2) 2024
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v19i2.27

Performance of the RMCNN4 Model
The RMCNN4 is the fourth model proposed. After it was implemented and tested using Fashion MNIST dataset, it produces the overall classification accuracy of 94.67%. Its visualization report as well as precision, Recall, and F1-Score are presented in Figure 8 and Table IV respectively.

Figure 8: Performance of the RMCNN4 Model

Table IV: The Performance of the RMCNN4 model

	Precision	Recall	F1-Score
T-shirt/Top	91%	82%	86%
Trousers	99%	99%	99%
Pullover	86%	88%	87%
Dress	91%	94%	93%
Coat	85%	89%	87%
Sandals	99%	99%	99%
Shirt	76%	76%	76%
Sneakers	97%	94%	96%
Bag	99%	98%	98%
Ankle boot	95%	98%	96%

Evaluation of the Proposed Model
The results of the proposed models were evaluated by first comparing the classification performance of the four proposed models. The model that produced the best classification accuracy (RMCNN3) was chosen as presented in Table V. It is therefore selected as the main model and compared with the existing models as well as the state of the arts in literature and conclusions were drawn based on its performance.

Table V: Accuracy of the proposed models

Model	Accuracy
RMCNN1	90.39%
RMCNN2	91.98%
RMCNN3	**95.81%**
RMCNN4	94.67%

The Proposed Model and the Existing Architectures.
The proposed model was compared with the base model as well as other existing architectures and it shows better performance as shown in Table VI

Table VI: The proposed Model with those of other Existing Architectures

Model	Accuracy
Lenet	90.16%
AlexNet	92.74%
ResNet	93.20%
MobileNet	93.96%
EfficientNet	93.64%
VIT	90.98%
MCNN15	94.04%
RMCNN3	**95.81%**

The Proposed Model and the Existing Models on Fashion MNIST Dataset
The proposed Model is now compared to those of other architectures trained using the Fashion MNIST dataset as presented in Table VII

Table VII: Proposed Model compared to other Models trained using Fashion MNIST Dataset.

.Model	Accuracy
H-CNN with VGG 16	94.00%
CNN-Softmax	91.86%
LSTMs	88.26%
LSTM	89.00%
SVM+HPO+REG	93.99%
CNNs	92.54%

CNN LeNet -5	90.64%
SVM + HOG	86.53%
CNN	89.54%
VGG	92.30%
Shallow CNN	93.69%
VGG Network	91.5%
MCNN15	94.04%
RMCNN3	**95.81%**

RESULT DISCUSSION
The performance of the proposed model is summarized in the Confusion metric as presented in Figure 9.

Figure 9: The confusion Matrix of the Proposed Model

Conclusion
In this research work, our focus was on developing a model that would reduce the effect of overfitting and improve the image classification accuracy of the Fashion MNIST dataset. This involves developing a simpler architectural model and adding a dropout layer which controls overfitting to every layer of the proposed model.
Subsequently, series of experiments were conducted to determine the optimal parameter values. The hyper-parameters were adjusted manually until the optimal results were obtained.
After comparing our proposed model with the base model, the state-of-the-art architectures, as well as other models built using Fashion MNIST dataset, it is very clear that the proposed model achieved an impressive classification accuracy of 95.81%, outperforming the existing models (by 1.77%)whose best result is 94.04%. This is probably due to the fact that dropout layer which controls overfitting were added to each of the network layers. Also, it is recorded in literature that simplified network architectures is one of the many ways used to control overfitting.
Although this research was able to improve the image classification performance of the Fashion MNIST dataset by about 1.77%, it is obvious that the proposed model took a larger amount of time for it to be properly trained. We, therefore, recommend that subsequent research works improve on the timing to produce a network model that would take less time.

REFERENCES
Baldi, P., & Sadowski, P. (2014). The dropout learning algorithm. *Artificial Intelligence*, *210*, 78–122. https://doi.org/10.1016/j.artint.2014.02.004
Bhatnagar, S., Ghosal, D., & Kolekar, M. H. (2017). Classification of fashion article images using convolutional neural networks. *2017 Fourth International Conference on Image Information Processing (ICIIP)*, 1–6. https://doi.org/10.1109/ICIIP.2017.8313740
Donati, L., Iotti, E., Mordonini, G., & Prati, A. (2019). Fashion Product Classification through Deep Learning and Computer Vision. *Applied Sciences*, *9*(7), 1385. https://doi.org/10.3390/app9071385
Duan, C., Yin, P., Zhi, Y., & Li, X. (n.d.). *Image Classification of Fashion-mnist Data Set Based on VGG Network.*
Eshwar S G, Gautham Ganesh Prabhu J, Rishikesh, A. V., Charan N A, & Umadevi V. (2016). Apparel classification using Convolutional Neural Networks. *2016 International Conference on ICT in Business Industry & Government (ICTBIG)*, 1–5. https://doi.org/10.1109/ICTBIG.2016.7892641
Greeshma, K. V., & Sreekumar, K. (2019). Fashion-MNIST classification based on HOG feature descriptor using SVM. *International Journal of Innovative Technology and Exploring Engineering*, *8*, 960–962.
Henrique, A. S., Fernandes, A. M. D. R., Lyra, R., Leithardt, V. R.

Q., Correia, S. D., Crocker, P., & Dazzi, R. L. S. (2021). Classifying Garments from Fashion-MNIST Dataset Through CNNs. *Advances in Science, Technology and Engineering Systems Journal*, *6*(1), 989–994. https://doi.org/10.25046/aj0601109

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., &Salakhutdinov, R. R. (2012). *Improving neural networks by preventing co-adaptation of feature detectors* (arXiv:1207.0580). arXiv. http://arxiv.org/abs/1207.0580

Lei, F., Liu, X., Dai, Q., & Ling, B. W.-K. (2020). Shallow convolutional neural network for image classification. *SN Applied Sciences*, *2*(1), 97. https://doi.org/10.1007/s42452-019-1903-4

Nocentini, O., Kim, J., Bashir, M. Z., & Cavallo, F. (2022). Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset. *Sensors*, *22*(23), 9544. https://doi.org/10.3390/s22239544

Ranjan, C. (2020). *Understanding Deep Learning Application in Rare Event Prediction*.

Saiharsha, B., Lesle A, A., Diwakar, B., R, K., & M, G. (2020). Evaluating Performance of Deep Learning Architectures for Image Classification. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 917–922. https://doi.org/10.1109/ICCES48766.2020.9137884

Seo, Y., & Shin, K. (2019). Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with Applications*, *116*, 328–339. https://doi.org/10.1016/j.eswa.2018.09.022

Shajini, M., & Ramanan, A. (2021). An improved landmark-driven and spatial–channel attentive convolutional neural network for fashion clothes classification. *The Visual Computer: International Journal of Computer Graphics*, *37*(6), 1517–1526. https://doi.org/10.1007/s00371-020-01885-7

Shajini, M., & Ramanan, A. (2022). A knowledge-sharing semi-supervised approach for fashion clothes classification and attribute prediction. *The Visual Computer*, *38*(11), 3551–3561. https://doi.org/10.1007/s00371-021-02178-3

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., &Salakhutdinov, R. (n.d.). *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*.

Su, H., Wang, P., Liu, L., Li, H., Li, Z., & Zhang, Y. (2020). *Where to Look and How to Describe: Fashion Image Retrieval with an Attentional Heterogeneous Bilinear Network* (arXiv:2010.13357). arXiv. http://arxiv.org/abs/2010.13357

Vijayaraj, A., asanth Raj, P. T., Jebakumar, R., Gururama Senthilvel, P., Kumar, N., Suresh Kumar, R., &Dhanagopal, R. (2022). Deep Learning Image Classification for Fashion Design. *Wireless Communications and Mobile Computing*, *2022*, 1–13. https://doi.org/10.1155/2022/7549397

Ying, X. (2019. An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022