JOURNAL OF SCIENCE, TECHNOLOGY AND ENGINEERING RESEARCH
Bilim, Teknoloji ve Mühendislik Araştırmaları Dergisi
ISSN (Online) 2717-8404
Available *online* at https://dergipark.org.tr/en/pub/jster

RESEARCH ARTICLE

Fourier-Based Image Classification Using CNN

[*]Göktuğ Erdem DAĞI,[1]Erhan GÖKÇAY, [2]Hakan TORA

[*]Atilim University, Faculty of Engineering, Electrical and Electronics Engineering Department, Ankara, Türkiye
goktugdagi@gmail.com, Orcid.0000-0001-5723-4578
[1]Atilim University, Faculty of Engineering, Software Engineering Department, Ankara, Türkiye
erhan.gokcay@atilim.edu.tr, Orcid.0000-0002-4220-199X
[2] Biruni University, Faculty of Engineering and Natural Sciences, Electrical and Electronics Engineering Department, Istanbul, Türkiye
htora@biruni.edu.tr, Orcid.0000-0002-0427-483X

Citation:
Daği, G.E., Gökçay, E., Tora H., (2024), Fourier-Based Image Classification Using CNN, *Journal of Science, Technology and* Engineering Research, 5(1):92-101. DOI: 10.53525/jster.1501920

HIGHLIGHTS

- *Processing of images in the frequency domain using CNN algorithm*
- *The available data is transformed into the frequency domain and fed to the CNN model.*
- *The proposed CNN model is trained by the spatial convolution.*
- *The high and low frequency components of the data introduces more useful properties than its raw pixels.*

Article Info

Received : June 15, 2024
Accepted : June 20, 2024

DOI:
10.53525/jster.1501920

***Corresponding Author:**

Göktuğ Erdem DAĞI,
goktugdagi@gmail.com

ABSTRACT

Recently, Convolutional Neural Networks (CNNs) have achieved remarkable success in computer vision, image processing and image processing tasks. Traditional CNN models work directly with spatial domain images. On the other hand, images obtained with Fast Fourier Transform (FFT) represent the Frequency domain and provide an advantage in computational cost by reducing potential calculation complexity. This study uses FFT converted images as input to the CNN algorithm to increase image classification and recognition accuracy and investigates the effects of this. The study begins with a comprehensive review of the foundations and features of FFT. It assumes that by converting the input images from the Spatial domain to the Frequency domain, the input image can be learned more efficiently and better results can be achieved in terms of performance by studying the most important features in the Frequency domain. To evaluate the effectiveness of this assumption, CIFAR-10, MNIST-Digits and MNIST-Fashion datasets were used. As a result, it has been shown that FFT-based preprocessing can improve classification accuracy, especially in cases where the datasets contain high-frequency noise, and it has shown different results in different datasets. Therefore, it is thought that the effect of FFT preprocessing varies depending on the datasets.

Keywords: *Machine Learning, Image Classification, Frequency Domain, Deep Learning*

I. INTRODUCTION

Artificial intelligence (AI) has rapidly developed in recent years and has made a significant impact in many industrial and academic fields. AI refers to the simulation of human intelligence and behavior in machines that are programmed to think and mimic human actions [1]. Advances in image processing and analysis have particularly highlighted the power and potential of AI [2]. Extracting meaningful information from image data has found applications in medicine, security, the automotive industry, retail, and many other fields [3]. These machines are designed to perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation [4]. In this context, the use of AI techniques, especially deep learning models like Convolutional Neural Networks (CNNs), has brought revolutionary advancements in the field of image processing [5]. AI systems can analyze vast amounts of data, recognize patterns, and make predictions or decisions based on this data [6]. Today, AI is one of the fastest-growing and most exciting fields in computer science [7]. AI aims to simulate human-like intelligence and behavior using computers, and it strives to achieve this goal through a series of algorithms and techniques [8]. One of the most important of these techniques includes artificial neural networks, a subfield known as deep learning [9]. Deep learning is achieved through the use of complex artificial neural networks trained on large amounts of data [10]. These networks simulate the process of learning from data and discovering complex patterns, similar to the way the human brain works [11]. One of the most effective and widely used types of deep learning models is known as Convolutional Neural Networks (CNNs) [3]. CNNs are a type of artificial neural network that excels particularly when working with image data [4]. Image processing can be defined as the area that allows a computer to interpret, understand, and process an image [5]. Image processing techniques are widely used in various fields such as object recognition, face recognition, medical imaging, autonomous vehicles, and security systems [6]. When combined with AI techniques, image processing becomes more complex and effective [7]. In this context, deep learning models such as CNNs show superior performance in various tasks in the field of image processing [8]. These advancements in image processing, particularly in overcoming the limitations of traditional pixel-based methods and developing new approaches for performing more complex tasks, have been significant.

There have been many techniques existing in the literature for image classification employing various CNN models. Akwaboah's study involved developing and training three different CNN models in the spatial domain to classify CIFAR-10 images. These models varied in their convolutional filter sizes, pooling layers, and the use of dropout regularization [12]. They achieved a test accuracy of 72.81%, 67.07, and 75.43 for the model 1, model2, and model3, respectively. Adeyinka's research focused on optimizing CIFAR-10 classification by testing various CNN models with different depths and configurations. Advanced architectures like ResNet and DenseNet were utilized, achieving test accuracies exceeding 90%. These models leverage deeper layers and skip connections, significantly enhancing model generalization and robustness [13]. Hengyue Pan proposed a novel approach by training a CNN model, named CEMNet, in the frequency domain. The approach simplifies the convolution operation, making it easier to parallelize by replacing it with element-wise multiplication [14]. CEMNet introduced several enhancements, including a weight fixation mechanism to mitigate overfitting, and adaptations of batch normalization, Leaky ReLU, and dropout layers for the frequency domain. Experimental results demonstrated that CEMNet could achieve over 70% accuracy on the CIFAR-10 dataset [14]. Classification was performed using two types of images: RGB images and Fourier-transformed images. Sophie Tötterström achieved an accuracy of 79.23% for RGB images and 38.78% for frequency domain images on the CIFAR-10 dataset [15].In another study, learnable frequency filters using Fourier transformation were developed for image classification problems. The main outlines of the study are as follows: Development of Frequency Filters, Cropping Procedure, and Weight Sharing. The developed learnable frequency filters achieved successful results in various computer vision problems. In summary, this study highlighted the significant contributions of frequency-domain approaches and learnable frequency filters in image classification problems. The developed methods demonstrated high performance in various application areas, making an important contribution to the literature [16].

This work presents a new approach that utilizes the frequency domain representation of the data to be classified in a CNN model. Instead of using frequency domain layers in the model, the available data is transformed to the frequency domain. In other words, we propose a convolutional network whose input is represented in the frequency domain. However, the model is trained by performing the spatial convolution.

The paper is organized as follows. In section 2, we describe the proposed approach and the data used for classifying the images. The experimental results and performance of the proposed method are introduced in section 3. Section 4 presents the conclusion.

II. MATERIAL AND METHOD

In this study, artificial neural networks (ANNs) inspired by the structure and function of the human brain were utilized. Artificial neural networks consist of interconnected nodes called neurons, which are organized in layers. Each neuron receives specific inputs, performs computations, and transmits its output to other neurons in the network. Artificial neural networks are trained using the backpropagation method. This method involves adjusting the internal parameters of the network (weights and biases) based on the error between the model's predictions and the actual target values. This iterative learning process enables neural networks to learn complex patterns and relationships in the data, making them effective for tasks such as classification, regression, and pattern recognition.

In this study, Convolutional Neural Networks (CNNs), a type of deep learning network particularly used for the analysis of visual data, were employed. CNNs are highly effective in tasks such as image recognition and classification. They automatically and adaptively learn spatial hierarchies from the input data.

Convolutional Neural Networks (CNN) Model

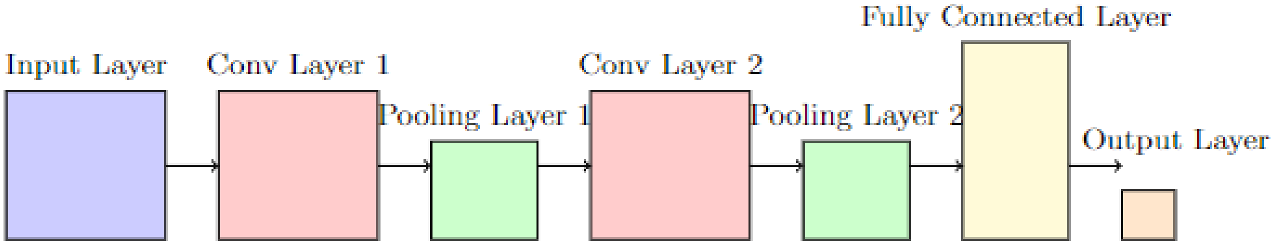


Figure 1: A simple representation of a Convolutional Neural Network(CNN) architecture.

The architecture of the CNN model consists of the following components:

- **Input Layer:** In CNNs, the input data typically consist of images or image-like datasets. This data is fed into the first layer of the CNN as matrices or tensors. For example, in RGB images, each pixel has three color channels: red, green, and blue. The input data are represented as matrices with pixel values ranging from 0 to 255.
- **Convolutional Layers:** Convolutional layers apply specific filters (kernels) to the input data to create feature maps. Filters are small matrices used to identify specific patterns or features. The convolution operation is performed by sliding these filters over the input data and calculating the dot product at each position to generate feature maps.
- **Activation Functions**: The outputs of the convolutional layers are usually subjected to a nonlinear activation function. The most commonly used activation function is the ReLU (Rectified Linear Unit) function. ReLU sets negative inputs to zero while keeping positive inputs unchanged, accelerating the learning process and reducing the vanishing gradient problem.
- **Pooling Layers:** Pooling layers are used to reduce the dimensions of the feature maps obtained from the

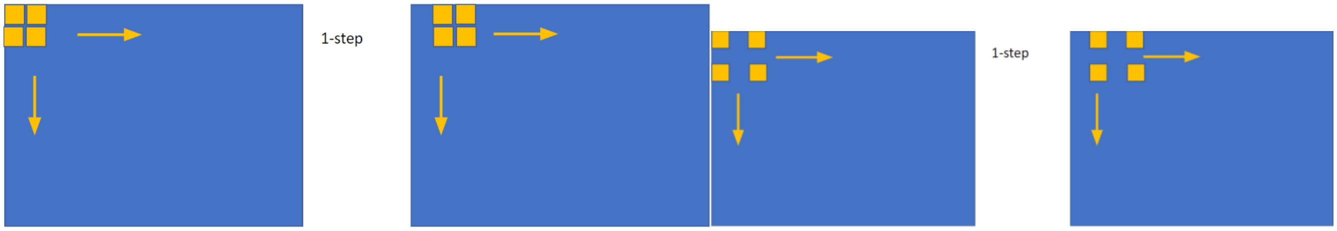
convolutional layers. This reduces the computational load of the model and mitigates the risk of overfitting. The most commonly used pooling type is max pooling, which selects the highest value in each region, thereby reducing the size of the feature map.

- **Fully Connected Layers:** These layers use the high-level features obtained from the convolutional and pooling layers to perform tasks such as classification or regression. The inputs are flattened and fed into fully connected layers, which use specific weights and biases to produce the final outputs.
- **Output Layer:** This layer produces the final predictions of the model. In classification tasks, the softmax activation function is used to obtain a probability distribution for each class.

CNN Training

The aim of this study was to investigate whether the image can be used by taking the Fourier Transform of the input images for image classification with convolutional neural networks. Models were created for both RGB images and Fourier spectra. Models trained and evaluated with RGB images were then compared with models trained via Fourier Transform. Additionally, various applications were made on the image while adjusting the data set. Datasets utilized include CIFAR-10, MNIST Fashion, and MNIST Digit. The CIFAR-10 dataset comprises 60,000 images distributed across 10 classes, with images sized at 32x32x3, denoting RGB color channels. Conversely, both MNIST Fashion and MNIST Digit datasets encompass 60,000 grayscale images at a size of 28x28. In the CNN model, preprocessing steps were meticulously applied to each 32x32x3 (RGB) image in the CIFAR-10 dataset with the aim of optimizing processing load and enhancing feature extraction capabilities.

Initially, the dataset was organized as a 70% training set, a 20% test set, and a 10% validation set..Then RGB images underwent a grayscale conversion, reducing them to a manageable 32x32 size to facilitate subsequent processing steps. Following this, each 32x32 image was divided into smaller 2x2 subimages. This subdivision process occurred across the entire row, with a stride value of 1, traversing from the zeroth index to the last index. Upon reaching the end of a row, the process was repeated in the subsequent row. This iterative process was performed across all rows and columns, resulting in a 50 percent overlap due to the stride value of 1. The purpose of this overlapping was to investigate potential correlations between adjacent data points shown in Figure 2.a. Additionally, 2x2 subimages were extracted again, this time with a shift process that left a gap between indexes to prevent overlap shown in Figure 2.b. Padding was applied throughout these operations to prevent loss of information at the leading and trailing indexes. This meticulous approach aimed to capture intricate details and patterns present in the images. Subsequently, a new dataset was generated by subjecting each of these small subimages to the Fourier Transform. This transformation enabled the extraction of frequency-based features from the images. Upon completion of the Fourier Transform, the dimensions of the new dataset were determined .To mitigate potential scaling issues, the data underwent scaling by applying the logarithm to each image. Following these preprocessing steps, the small 2x2 subimages were reassembled to restore them to their original 32x32 dimensions shown in Figure 3. This reconstruction process ensured that the fundamental features extracted from the images remained intact. These comprehensive preprocessing steps played a pivotal role in enhancing the model’s capability to extract meaningful features from the CIFAR-10 dataset, thereby contributing to the overall improvement in the classification performance of the CNN model. Furthermore, all operations except grayscale conversion were applied to the MNIST-Digits and MNIST-Fashion datasets, ensuring consistency in preprocessing across different datasets.



(a) with overlap

(b)non-overlap

Figure 2: (a) refers to the overlapping image and (b) refers to the non-overlapping image.

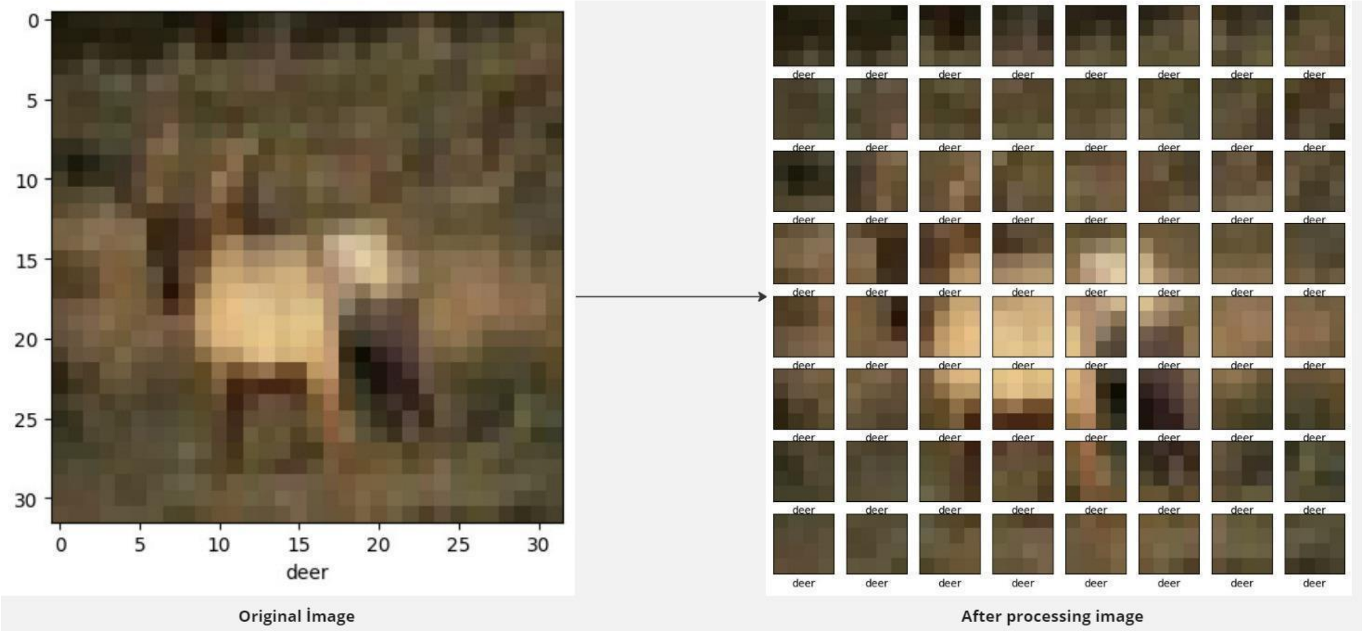


Figure 3: Original image and after processing.

Proposed Model

In this study, the CNN algorithm is used for image classification. The proposed model is illustred in Figure 4.The algorithm consists of 4 layers. The kernel size is set to 3x3, ”same” padding is applied, and ”ReLu” activation function is used. Additionally, BatchNormalization, Dropout, and MaxPooling are employed. In the fully connected layer, the ”softmax function” is used as the activation function. ”Sparse categorical crossentropy” is utilized as the loss function, and ”Adam” is chosen as the optimizer.

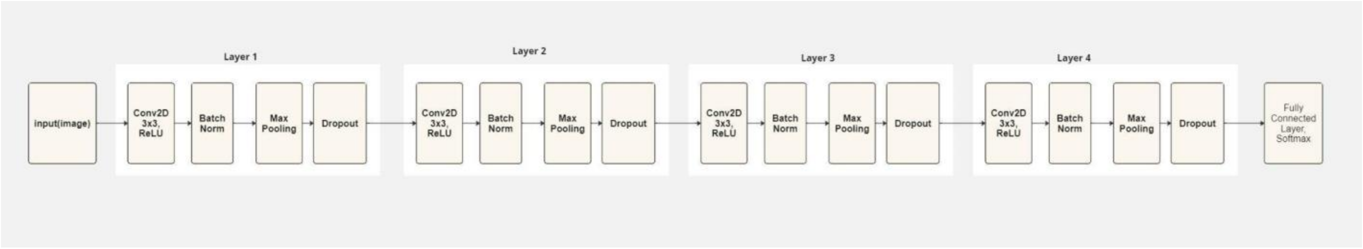


Figure 4: Proposed Model

Application of the Basic Model and Training Process

Following the application of the basic model, a classification model specific to the application studied was developed based on the same basic model. In this context, four different models were trained for the CIFAR-10, MNIST-Digits, and MNIST-Fashion datasets. For the CIFAR-10 dataset, one of the trained models performed the sliding operation with a standard stride value of 1, while the other models performed the sliding operation with 50% and 0% overlap, respectively. The same processes were repeated for the MNIST-Digits and MNIST-Fashion datasets. The trained models will provide benchmark results for the experiments conducted. Figures 5, 6, and 7 show the Training and Validation Accuracy values and Training and Validation Loss values for the CNN model throughout the training process on the CIFAR-10,MNIST-Digits, and MNIST-Fashion datasets. These figures indicate that both Training and Validation Loss values decreased, suggesting that the model effectively learned the data.

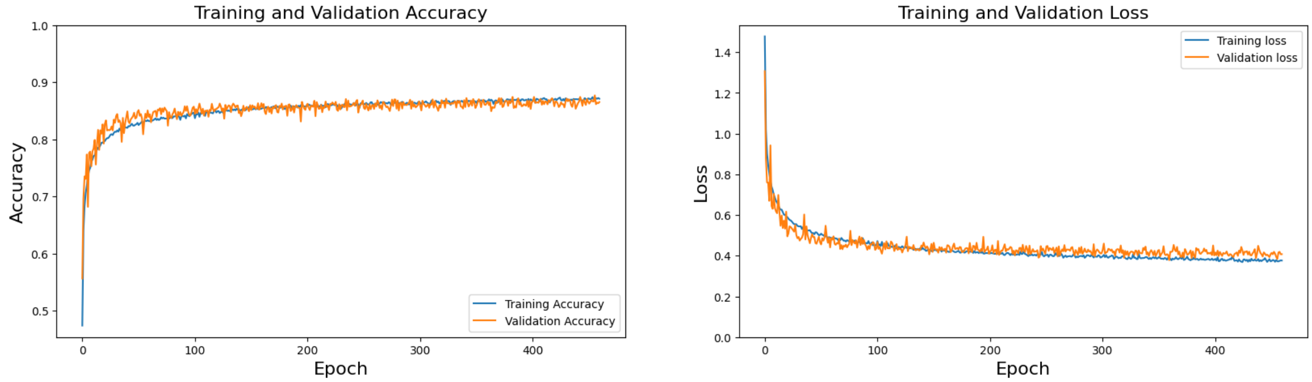


Figure 5: Training and Validation results for CIFAR-10 dataset

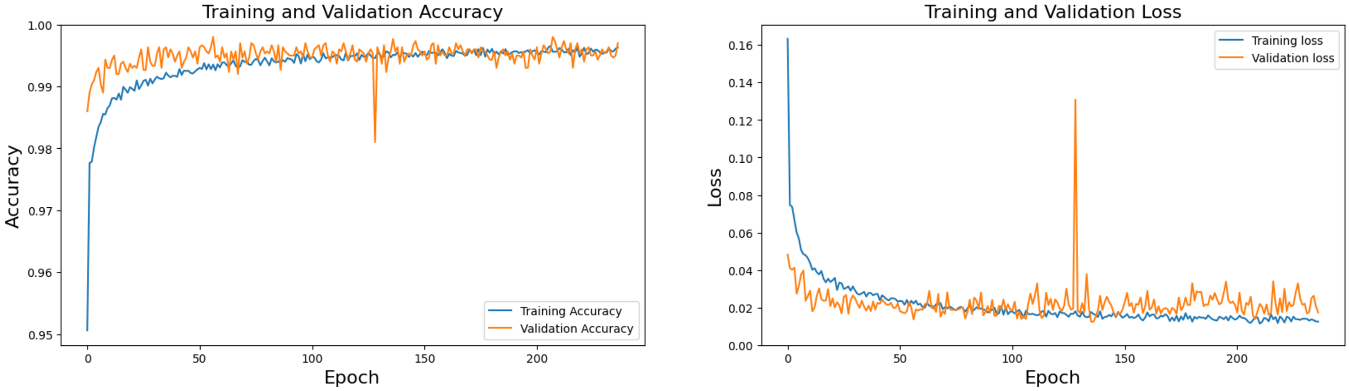


Figure 6: Training and Validation results for MNIST-Digits dataset

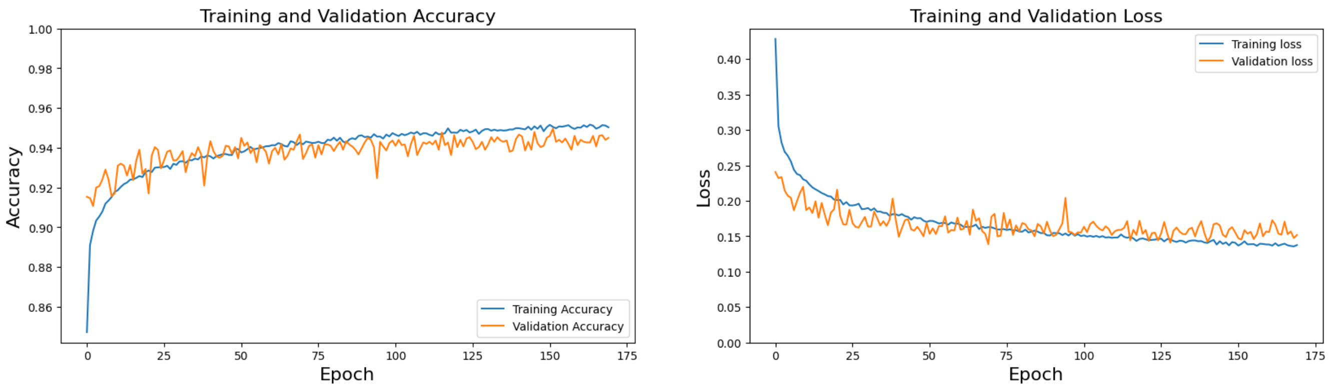


Figure 7: Training and Validation results for MNIST-Fashion dataset

Figures 8, 9, and 10 present the results of the Fourier Transform. These figures also show that, despite some occasional spikes, both Training and Validation Loss values generally decreased. Similarly, the Training and Validation Accuracy values were observed to increase. These findings demonstrate that the model effectively learned.

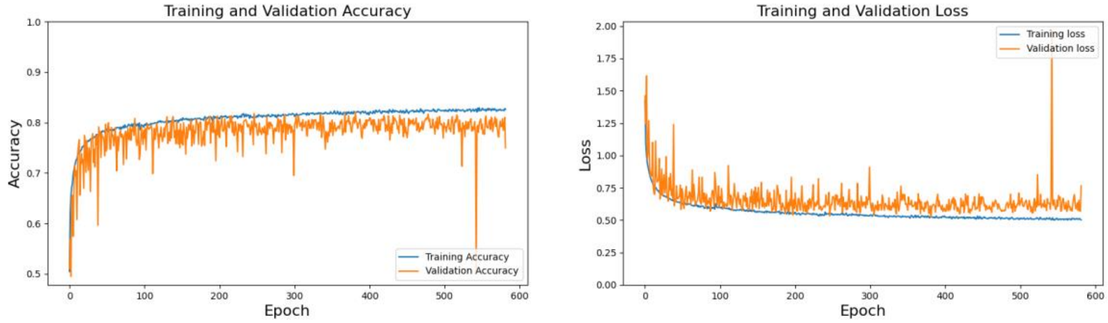


Figure 8 : Fourier Transform results for CIFAR-10 dataset

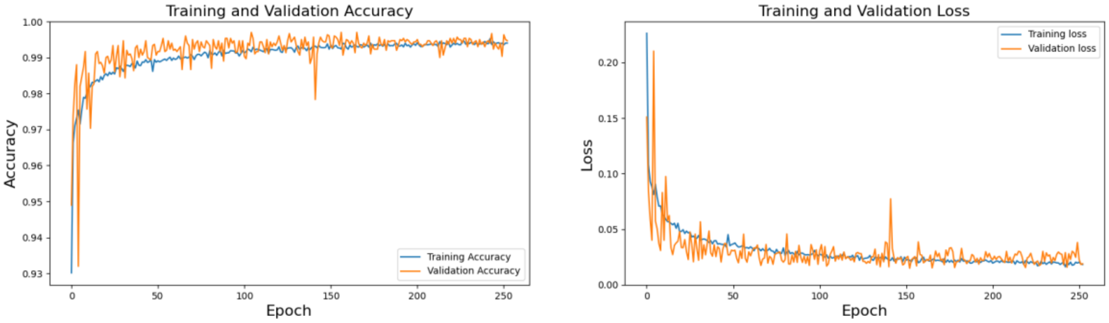


Figure 9 : Fourier Transform results for MNIST-Digits dataset

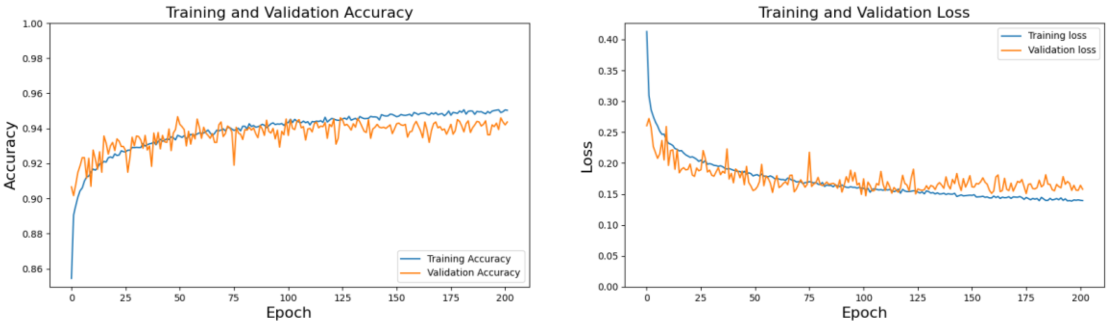


Figure 10 : Fourier Transform results for MNIST-Fashion dataset

III. RESULTS AND ANAYSIS

This study is designed to examine the effectiveness of CNN models in image classification in the frequency domain. Model architectures and training processes were evaluated on various datasets such as CIFAR-10, MNIST-Digits, and MNIST-Fashion. The loss values used during training indicate how reliably the model adapts to the data. The loss values obtained during training in Figures 5, 6, and 7 demonstrate that the selected model adapts more reliably to the data. However, this reliability does not reflect in the classification results on the test data. The accuracy values in Table 1 indicate that the model's ability to classify test data is weaker than the model discussed in the study. CNNs were observed to be more suitable for image classification in the time domain. This model architecture achieved better performance in image classification in the time domain. It was observed that CNNs were more suitable for image classification in the time domain. The results obtained from classifying input images obtained through Fourier transformation did not outperform those obtained in the time domain classification. However, it was observed that the proposed model was more successful compared to the operations performed on the entire image. The observed results are given in Table 1.This indicates that Fourier transformation may lead to information loss in some cases and adversely affect classification performance. In conclusion, this study evaluated the effectiveness of CNN models in image classification in the time domain. The results demonstrate that image classification in the time domain is more successful compared to frequency domain approaches such as Fourier transformation. However, more comprehensive studies on different datasets and model architectures will contribute to a deeper understanding of these results.

Table 1 : Accuracy values for different datasets and domains

Dataset	Spatial Domain	Frequency Domain without subimages	Frequency Domain with subimages
CIFAR-10	0.8593	0.4771	0.7865
MNIST-Digits	0.9946	0.9418	0.9326
MNIST-Fashion	0.9341	0.8362	0.8752

IV. CONCLUSION

This paper explored the effectiveness of utilizing Fast Fourier Transform (FFT)-transformed images as inputs for Convolutional Neural Networks (CNNs) in image classification tasks. Various CNN architectures and training methodologies were assessed using benchmark datasets, including CIFAR-10, MNIST-Digits, and MNIST-Fashion. Training loss values, as depicted in Figures 6, 7, and 8, showed that the models adapted well to the data. However, this adaptation did not necessarily result in improved classification accuracy on the test data, as evidenced by the accuracy values presented in Table 1.The research revealed that CNNs performed more effectively in the spatial (time) domain compared to the frequency domain. However, performing operations as in the proposed model instead of all images in the frequency domain gave better results. Models trained on spatial domain images consistently achieved better results than those trained on FFT-transformed images. This suggests that information necessary for accurate classification may be lost during the Fourier transform, negatively affecting the performance of the model. Although the frequency domain approach offered some valuable insights into data representation, it did not surpass the traditional spatial domain approach in terms of results. This underscores the robustness of spatial domain data for CNN-based image classification tasks. The observed performance gap between the two domains indicates that further refinement and the development of hybrid approaches might be required to fully harness frequency domain information. In conclusion, this study demonstrates that image classification with CNNs is more successful in the spatial domain than in the frequency domain. While FFT provides an alternative perspective on image data, it may introduce challenges that can hinder classification performance. Future research should focus on developing hybrid models that incorporate both spatial and frequency domain information, as well as examining the effects of FFT on a broader range of datasets and more advanced CNN architectures. Such efforts will enhance our understanding of the strengths and limitations of frequency domain analysis in deep learning. The findings of this work contribute to the field of computer vision by providing insights into the use of frequency domain techniques in CNN workflows. Future studies are encouraged to validate these findings and explore innovative methodologies to improve image classification performance by effectively combining spatial and frequency domain data.

CONFLICTS OF INTEREST

They reported that there was no conflict of interest between the authors and their respective institutions.

RESEARCH AND PUBLICATION ETHICS

In the studies carried out within the scope of this article, the rules of research and publication ethics were followed.

REFERENCES

[1] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach," 3rd ed., Prentice Hall, Upper Saddle River, NJ, USA, 2020.

[2] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," MIT Press, Cambridge, MA, USA, 2016.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097-1105.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 770-778.

[5] C. Szegedy et al., "Going Deeper with Convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2015, pp. 1-9.

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. Int. Conf. Learning Representations*, 2014.

[7] V. Mnih et al., "Playing Atari with Deep Reinforcement Learning," in *Advances in Neural Information Processing Systems 27*, 2013, pp. 1-9.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL-HLT*, 2018, pp. 4171-4186.

[9] I. Goodfellow et al., "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems 27*, 2014, pp. 2672-2680.

[10] A. Vaswani et al., "Attention is All You Need," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 5998-6008.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "R-CNN: Regions with Convolutional Neural Network Features," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 580-587.

[12] Akwasi Darkwah Akwaboah, "Implementation of Convolutional Neural Networks for CIFAR-10 Image Classification," 2019.

[13] Ajala Sunday Adeyinka, "Convolutional Neural Network Implementation for Classification using CIFAR-10," *ResearchGate*, 2023.

[14] Hengyue Pan, "Learning Convolutional Neural Networks in Frequency Domain," *ResearchGate*, 2023.

[15] S. Tötterström, "Frequency Domain Image Classification with Convolutional Neural Networks," *Bachelor's Thesis, Tampere University*, 2023.

[16] Stuchi, J. A., Canto, N. G., de Faissol Attux, R. R., & Boccato, L. (2024). A frequency-domain approach with learnable filters for image classification. *Applied Soft Computing, 155*, 111443.