Deep learning methods for skin cancer classification | MNIST dataset

Artem Khomytskyi - 20221686 | Timofii Kuzmenko - 20221690 | Davyd Azarov - 20221688 | Luís Soeiro: 20211536

General Idea

Skin cancer is the out-of-control growth of abnormal cells in the epidermis, the outermost skin layer, caused by unrepaired DNA damage that triggers mutations. These mutations lead the skin cells to multiply rapidly and form malignant tumours. Statistically, 1 out of 5 Americans will develop skin cancer by age 70. [1]. The project consists of two stages: in the first one, we are required to develop a binary classification model to differentiate between male and female patients based on medical images. Biological and lifestyle factors influence how skin cancer manifests, with males having thicker, coarser skin and more irregular, darker lesions, often appearing on the scalp, ears, and shoulders due to greater sun exposure and delayed diagnosis. In contrast, females have thinner skin, hormonally influenced pigmentation, and more frequent lesions on the legs and arms due to different sun exposure patterns.

The study focuses on two key objectives: binary classification of patient gender and multi-class classification of specific lesion types. Preprocessing steps included resizing images to 128×128 pixels, normalization, and data augmentation to address variability and class imbalance.This time we were using the MNIST dataset, which consists of a large collection of skin lesion images designed for machine learning research in dermatology. This particular dataset was released as part of the **Human Against Machine (HAM10000) challenge** [4]. The model analyzes lesion texture, size, and location to predict sex, but further validation is needed to ensure it identifies biological differences rather than dataset biases. In the second stage, we extended the model to perform multi-class classification, identifying 7 specific tumour types. As a result of our work, DenseNet121 (accuracy 0.49) and Baseline CNN (accuracy 0.64) showed better performance for multiclass classification depending on the target column, while FCNN with ReLU and Tanh performed best for binary classification (accuracy 0.55), according to the study's evaluation of several deep learning models for binary and multiclass classification tasks on dermatoscopic images. These findings demonstrate how crucial transfer learning and convolutional architectures are for efficiently identifying spatial patterns and handling challenging classification problems. In Future Work, we will focus on improving the architecture and depth of the model relative to our hardware, while maintaining computational efficiency, to further improve classification accuracy.

KEYWORDS - Skin Cancer Diagnosis, Deep Learning, Convolutional Neural Networks, Transfer Learning, Medical Imaging, Image Preprocessing

Introduction

The most common cancer in the United States and worldwide is skin cancer, having depressing statistics of More than 2 people dying of skin cancer in the U.S. every hour. Having 5 or more sunburns doubles your risk for melanoma. More than 5.4 million cases of nonmelanoma skin cancer were treated in over 3.3 million people in the U.S. in 2012, still considered the best estimate to date. Chances are, you know at least one person who has been personally affected by skin cancer [5]. When caught in its earliest, localized stages, the 5-year relative survival rate for melanoma is 99%. Advances in early detection and treatment methods have significantly increased skin cancer survival rates in recent years, and the data from 2014 – 2020 shows that across all stages of melanoma, the average five-year survival rate in the U.S. is 94%. [5]. Thus, ML and DL models have the potential to be a decisive argument in human survival. With the improvement of the quality of models, it becomes possible to detect pathology at ever earlier stages and more accurately.

There are several traditional techniques for skin cancer diagnosis, such as dermatoscopy (a non-invasive technique where a special handheld device is used to magnify and illuminate the skin lesion to identify characteristic features of skin cancer), ultrasound, and biopsy(taking a small sample of cells) [6], that play a critical role in skin cancer diagnosis. However, histopathological image analysis is still used as one of the most definitive methods for detecting and classifying tumours. It requires pathologists to identify patterns and textures to indicate malignancy. Computer-Aided Detection (CAD) [7] are systems that assist doctors in the interpretation of medical images. Imaging techniques in X-ray, MRI, endoscopy, and ultrasound diagnostics yield a great deal of information that the radiologist or other medical professional has to analyze and evaluate comprehensively in a short time. However, so far its application has been limited to quantifying immunostaining. Recent advances in machine learning (ML) and deep learning (DL) enable us to create more accurate and automated diagnostic tools in medical imaging.

Deep Learning, particularly Convolutional Neural Networks (CNNs), has shown success in medical image classification, processing large datasets and identifying intricate patterns. The deep learning revolution of the 2010s has already produced AI that is more accurate in many areas of visual diagnosis than radiologists and dermatologists, and this gap is expected to grow [8]. As DL models, in general, are being applied to almost every part of everyday life, models with a similar structure as the models we use in our project are used in such medical fields as dermatology, neurology, and cardiology, showing promising results in detecting skin cancer, brain tumours, and cardiovascular diseases. The findings showcase the CNN model's exceptional performance in distinguishing cancerous and non-cancerous skin tissue in the integrated dataset, thereby demonstrating its potential for enhancing clinical decision-making and fostering the development of AI-driven diagnostic solutions [9]. Transfer learning techniques, which utilize pre-trained models, have further enhanced performance by adapting deep learning models to medical imaging tasks.

Despite all the previous being said, some challenges remain in implementing DL models for medical diagnostics. When it comes to scaling the performance of the models, it often requires extensive computational resources and even larger datasets than before, which can be limited due to

ethical concerns regarding medical data. In addition, deep learning models are often associated with “black boxes,” meaning that it is difficult to interpret how the model makes decisions, which may raise concerns about their clinical applicability among the general public. To avoid such problems, it is necessary to integrate more transparent AI models, improve the availability of datasets, and continuously improve neural network architectures.

In this project, we aimed to develop deep-learning models capable of detecting and classifying skin cancer using dermatoscopic images from the MNIST dataset. The final dataset consists of 10015 dermatoscopic images. Cases include a representative collection of all important diagnostic categories in the realm of pigmented lesions: Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv) and vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and haemorrhage, Vasc) [4]. Our main objectives were to build and further optimize CNN-based models for two key classification tasks: binary classification to distinguish between male and female patients and multi-class classification to identify the specific tumour type. During our work, we explored various image preprocessing techniques, experimented with different CNN architectures, and applied transfer learning to enhance model performance. We believe that by using deep learning for automated skin cancer diagnosis, we contribute to the ongoing efforts to develop efficient, accurate, and scalable diagnostic tools in medical imaging.

Data Exploration

For this project we have used a dataset that originates from the HAM10000 database, initially containing 10,015 dermoscopic images. During initial preprocessing, 391 duplicate/corrupted images were removed, leaving us with a final dataset of 9,624 samples. These samples represent seven diagnostic categories of skin cancer, which are: melanocytic nevi (nv), melanoma (mel), benign keratosis-like lesions (bkl), basal cell carcinoma (bcc), actinic keratoses (akiec), vascular lesions (vasc), and dermatofibroma (df). The dataset is unbalanced, with the largest class (nv) representing approximately 70% of the images, while the smallest class (df) accounts for less than 2%.

The metadata associated with the given images included patient age, gender, and lesion localization. Some columns were not directly related to our area of research (age and lesion_id) therefore were dropped. The images were originally in JPEG format with predominantly resolutions of 600x450 pixels.

The dataset was split into three subsets: training (70%), validation (15%), and testing (15%). This approach ensures that class imbalance does not disproportionately affect any subset, thereby supporting robust model performance evaluation.

The images were organized into directories corresponding to their diagnostic classes and split categories. This hierarchical structure allows for efficient image retrieval during the training and testing phases.

Visual inspection of the dataset revealed significant differences between cancer categories. For example, melanocytic nevi (nv) generally have uniform pigmentation, while melanoma (mel) lesions have irregular borders and varying color patterns.

Challenges and Considerations

Because minority classes like df and vasc were underrepresented, the dataset's inherent class imbalance made model training difficult. This was addressed by artificially increasing the representation of these classes in the training set using methods like data augmentation. Furthermore, effort was made to prevent artifacts from being introduced by augmented samples, which could skew model predictions.

Conclusion and Next Steps:Implementing and optimizing deep learning models to categorize lesions into the appropriate diagnostic groups will be the main focus of future research. Addressing class imbalance and utilizing cutting-edge feature extraction techniques to increase model fidelity will be the main priorities.

Data Preprocessing

Image Resizing and Input Dimensions: To standardize the input images for the deep learning model, all images were resized to a resolution of 128x128 pixels. This resolution was chosen to balance computational efficiency and detail preservation, ensuring compatibility with memory constraints and model input requirements.

Label Encoding: The categorical labels associated with the images were encoded using one-hot encoding or Sklearn’s LabelEncoder. This encoding transformed the labels into numerical values, making them suitable for input into machine learning algorithms. For multiclass classification problems, the labels were converted to class indices after encoding.

Handling Class Imbalance: To address class imbalance in the dataset, we emplied resampling techniques such as RandomOverSampler or RandomUnderSampler. Oversampling involved duplicating minority class samples, while undersampling reduced the majority class samples, resulting in a balanced dataset. This ensured that the model did not become biased towards the majority class during training.

Normalization and Tensor Conversion: To improve generalization and reduce overfitting, data augmentation was applied to the training set using ImageDataGenerator from Keras and TensorFlow's Image Processing API. The augmentation techniques used included: **- Rotation:** Random rotations of up to 20 degrees. **- Width and Height Shifting:** Shifting images by up to 10%. **- Zooming:** Random zoom-in and zoom-out

transformations. **- Shearing:** Geometric distortions to introduce variation. **- Horizontal Flipping:** Mirroring images to simulate diverse perspectives.

Normalization and Tensor Conversion: Pixel values of all images were normalized to the range [0, 1], as neural networks typically perform better with scaled inputs. Additionally, the image arrays were converted into tensors.

Dataset Creation: The preprocessed data was transformed into TensorFlow datasets using tf.data.Dataset. These datasets were further optimized with: **- Shuffling:** Applied to the training dataset to ensure random sampling of batches during training. **- Batching:** The datasets were batched with a size of 32 images per batch. **- Prefetching:** Utilized TensorFlow's AUTOTUNE setting to overlap data preprocessing and model execution.

Class Weights: During training, class weights were calculated to compensate for any residual imbalance. This step helped mitigate the effect of uneven sample distribution on training performance.

Methodology

Setup for implementation

This project was developed using the **HAM10000 dataset**, a collection of **10,015 dermatoscopic images** categorized into seven skin lesion types. Due to the computational demands of deep learning, models were trained using **Visual Studio Code**. The main computing environment included **MacOS Monterey, Quad-Core Intel Core i5, Intel Iris Plus Graphics 645, 8GB RAM, Python 3.11.8, and TensorFlow 2.10.0** (for compatibility with GPU support). Transfer learning and deep neural networks require significant computational power, making GPU acceleration essential for faster convergence.

Model Development for Binary Classification

Using fully connected networks (FCNNs), we investigated three methods for binary categorisation of medical pictures. Our main objective was to use deep learning approaches to differentiate between two classes (male vs. female, for example) while addressing overfitting and optimising feature learning.

1. FCNN With ReLU and Dropout: The first model was created from the ground up to use dense layers to learn patterns. Before putting the input images through two dense layers (256 and 128 neurones) using ReLU activation to add non-linearity, it flattened them into a 1D vector. After every layer, batch normalisation was used to stabilise training, and dropout (rate = 0.3) assisted in reducing overfitting. Lastly, probability scores for the binary classification were obtained from a sigmoid output. The model had trouble capturing the subtleties of space present in visual data, even if it was computationally efficient.

2. FCNN With Tanh and ReLU Activation: This network improved on the first method by adding a tanh activation in the first dense layer (512 neurones) and ReLU in the layers that followed (256 and 128 neurones). Additionally, dropout (rate = 0.4) and batch normalisation were used to enhance generalisation and convergence. Tanh may have improved feature learning by permitting both positive and negative activation ranges. Even though accuracy was higher than the first model, precise spatial feature extraction was still constrained by the lack of convolutional filters.

3. FCNN With L2 Regularization and Higher Dropout: Stronger regularisation was the main emphasis of the third model, which included dropout (rate = 0.5) and L2 (weight decay) in each dense layer (512, 256, and 128 neurones) to further prevent overfitting. After ReLU activation, each layer underwent batch normalisation, which stabilised weight updates. Even while this model was more resilient to overfitting than the previous two, it was still fundamentally limited by the absence of convolutional layers for the extraction of spatial features.

Model Development for Multiclass Classification

Using skin lesion photos, we investigated three sophisticated methods for multiclass: DenseNet121, a transfer-learning-based ResNet50, and a baseline Convolutional Neural Network (CNN). The main goal was to use deep learning techniques to accurately classify lesions into numerous classifications.

1. CNN baseline: In order to understand hierarchical spatial patterns, the baseline CNN was created from the ground up. It included max-pooling layers to downsample feature maps and convolutional layers with progressively larger filter sizes (32, 64, and 118). A softmax activation function was used to link learnt features to output probabilities through the addition of fully connected dense layers. To reduce overfitting, dropout layers (rate = 0.3) were added, and the model was assembled using the Adam optimiser with categorical cross-entropy loss. The CNN's restricted depth made it difficult for it to generalise on complex lesions, even with respectable performance.

2.VGG 16: We used VGG16, a deep convolutional network pre-trained on ImageNet, to enhance feature extraction and shorten training time. Thirteen convolutional layers and fully connected layers made up the design, which used tiny 3x3 filters for hierarchical feature learning. We added global average pooling, dropout, and a softmax activation function to improve the classification layers after freezing the original convolutional layers to protect previously learnt information. While using fewer training epochs, this transfer learning method greatly increased accuracy when compared to the baseline CNN. However, VGG16 was computationally costly and prone to overfitting without adequate regularisation because of its huge number of parameters.

3. ResNet50: To enhance feature extraction, we used a transfer learning approach with ResNet50 pre-trained on ImageNet. The model's lower layers were frozen to retain pre-trained features, while custom classification layers, including a global average pooling layer, dense layers with ReLU activation, and a softmax output, were fine-tuned for our dataset. This approach significantly reduced training time and improved performance by leveraging knowledge from large-scale image datasets.

4.Fully Connected Networks (without using CNN): We developed a model that flattened picture input into a 1D vector before putting it through several dense layers in order to evaluate how well a conventional fully connected neural network (FCNN) performed image classification. Three dense layers (512, 256, and 128 neurones) with batch normalisation, dropout (rate = 0.4) for regularisation, and ReLU activation were part of the architecture. Despite their computational efficiency, FCNNs were far less accurate than CNN-based models because they were unable to capture spatial hierarchies. The model had a wide discrepancy between training and validation accuracy due to overfitting and high variation.

5.Autoencoder and Classifier: We used an Autoencoder + Classifier strategy to investigate unsupervised pre-training, which includes: An autoencoder to learn compressed latent representations of pictures. These encoded representations are used to train a classifier network for final classification. A CNN-like structure was used for the encoder component, which first captured pertinent characteristics before lowering dimensionality. By rebuilding input images, the decoder made sure that feature learning was meaningful. The autoencoder outperformed CNN-based models in classification, especially when it came to differentiating between visually similar lesion types, even though it was better at feature extraction than FCNNs.

6. DenseNet121: Finally, we implemented DenseNet121, known for its dense connectivity, which promotes feature reuse and alleviates vanishing gradient issues. Using pre-trained ImageNet weights, the model was fine-tuned with custom layers similar to ResNet50 but benefited from its more efficient gradient flow and feature propagation.

Data Preparation: To improve the robustness of each model, we carried out data preprocessing, which included scaling, normalisation, and data augmentation (rotations, flips, and zooming). We used oversampling techniques to ensure that all classes were fairly represented in order to alleviate the class imbalance. Evaluation and Findings: Confusion matrices were used to get insight into the performance of each class, and models were assessed using metrics such as accuracy, precision, recall, and F1-score. ResNet50 demonstrated a significant increase in accuracy and generalisation, whilst the baseline CNN offered fundamental insights. But DenseNet121 was the best model for multiclass classification in our study, outperforming both by obtaining the highest accuracy and F1-scores.

Results

Model	Column Observed	Class	Evaluation Metrics						
			Accuracy	Macro Avg.			Weighted Avg.		
				Precision	Recall	F1-Score	Precision	Recall	F1-score
FCNN ReLU	sex	Binary	0.50	0.55	0.53	0.45	0.55	0.50	0.44
FCNN ReLU and Tanh	sex	Binary	0.55	0.27	0.50	0.35	0.30	0.55	0.39
FCNN L2	sex	Binary	0.46	0.56	0.50	0.31	0.57	0.46	0.29
Baseline CNN	dx	Multiclass	0.64	0.75	0.73	0.73	0.75	0.73	0.73
VGG16	dx	Multiclass	0.52	0.55	0.51	0.48	0.55	0.51	0.48
ResNet50	dx_type	Multiclass	0.41	0.42	0.50	0.43	0.42	0.50	0.53
FCNN ReLU	dx_type	Multiclass	0.25	0.06	0.25	0.10	0.06	0.25	0.10
Classifier	localization	Multiclass	0.39	0.24	0.28	0.22	0.24	0.28	0.22
DenseNet121	localization	Multiclass	0.49	0.41	0.42	0.40	0.41	0.42	0.40

Stronger regularisation was the main emphasis of the third model, which included dropout (rate = 0.5) and L2 (weight decay) in each dense layer (512, 256, and 128 neurones) to further prevent overfitting. After ReLU activation, each layer underwent batch normalisation, which stabilised weight updates. Even while this model was more resilient to overfitting than the previous two, it was still fundamentally limited by the absence of convolutional layers for the extraction of spatial features.

Models were assessed across various target columns ("dx," "dx_type," and "localisation") in multiclass classification. The baseline CNN outperformed VGG16 (accuracy 0.52) in terms of generalisation, achieving the maximum accuracy (0.64) and F1-score (0.73) for the "dx" column. But when it came to "localisation" classification, DenseNet121 outperformed the others, attaining the best accuracy (0.49) and balanced evaluation metrics (Macro F1-score 0.40). When compared to the FCNN-based model, ResNet50 performed marginally better for "dx_type" (accuracy 0.41), demonstrating the benefits of transfer learning in utilising pre-trained feature hierarchies. These findings highlight how a model's performance varies depending on the target column, with DenseNet121 and ResNet50 demonstrating resilience in more challenging multi-class tasks.

Conclusion

This work investigates deep learning methods for classifying skin cancer using the HAM10000 dataset. The study presents two challenges: (1) the multi-class classification of seven types of skin lesions and their localisation, and (2) the binary classification of patient gender based on lesion characteristics. The 10,015 photos in the dataset were reduced to 9,624 images following preprocessing. Normalisation, picture scaling to 128×128 pixels, and substantial data augmentation were important preprocessing methods to address variability and class imbalance. For binary classification, Fully Connected Neural Networks (FCNNs) were assessed using various regularisation techniques and activation functions. "FCNN ReLU and Tanh," the top binary model, had a 55% recall, an F1-score of 0.39, and an accuracy of 55%. These models' performance was constrained by their incapacity to process spatial image data, even if they were able to capture fundamental patterns. For multi-class classification, the baseline CNN achieved 64% accuracy and a 0.73 F1-score but struggled with complex lesion patterns. Transfer learning models significantly improved performance. VGG16 achieved 52% accuracy and a 0.48 F1-score, while ResNet50 achieved 41% accuracy and a 0.53 F1-score. DenseNet121 was the best-performing model, achieving 49% accuracy and a 0.42 F1-score, highlighting the advantages of dense connectivity and pre-trained feature extraction for medical imaging. The findings highlight the potential of deep learning, particularly transfer learning, in enhancing the accuracy and reliability of skin cancer classification models. This research demonstrated that leveraging advanced architectures like DenseNet121 can improve feature learning and classification performance, even with imbalanced datasets. This research underscores the importance of deep learning in developing AI-driven diagnostic tools, offering a promising step toward improving early detection and treatment of skin cancer.

Future Work

By using more neural networks, hyperparameters, and attempting using more complicated transfer learning methods, our future work will attempt to improve the model's structure and performance. In order to address the issues of data scarcity in medical imaging, we will also test self-supervised and semi-supervised learning techniques to lessen reliance on sizable labelled datasets. Enhancing model interpretability using explainable AI methods will be our top priority if this project is expanded, guaranteeing that physicians can trust the model's judgements. Last but not least, gaining more data that will provide us with a wider range of demographics and novel characteristics, such as genetic markers and patient history, may improve diagnostic precision and generalisation even more.

References

[1] **The Skin Cancer Foundation.** "Skin Cancer Information," 2025. [Online].

[2] **M. Bellenghi et al.** "Sex and Gender Disparities in Melanoma," *Cancers*, vol. 12, no. 7, p. 1819, Jul. 2020. [Online]

[3] **A. C. Baumann et al.** "Sex Disparities in Cutaneous Melanoma Outcomes," *JAMA Dermatology*, vol. 156, no. 5, pp. 553–560, May 2020. [Online]

[4] **Harvard Dataverse.** "GLOBOCAN 2020: Estimated Cancer Incidence, Mortality, and Prevalence Worldwide." *Harvard Dataverse*, 2020. [Link].

[5] **The Skin Cancer Foundation.** "Skin Cancer Facts & Statistics," 2025. [Online].

[6] **Winship Cancer Institute of Emory University.** "Skin Cancer Diagnosis," 2025. [Online]

[7] **"Computer-aided Diagnosis: The Tipping Point for Digital Pathology".** Digital Pathology Association. 27 April 2017.

[8] Paiva, Omir Antunes; Prevedello, Luciano M. (October 2017). **"The potential impact of artificial intelligence in radiology"**

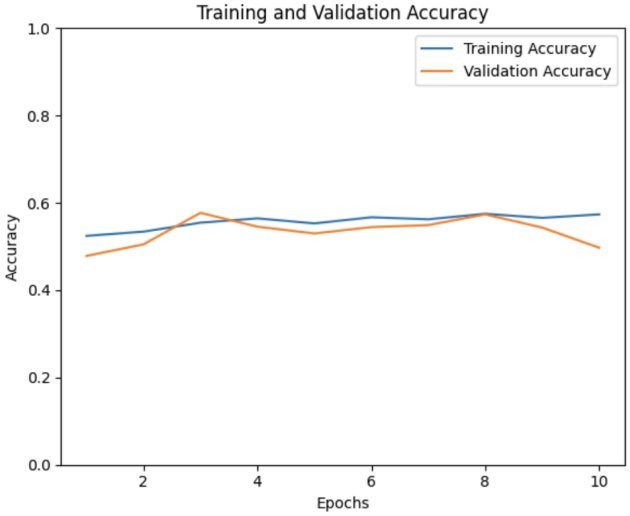
[9] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning," Healthcare, vol. 10, no. 7, p. 1183, Jun. 2022. [Online].

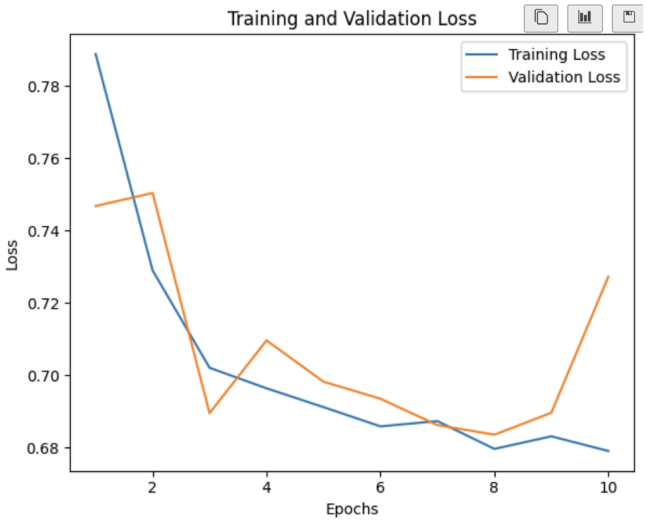
Additional Information

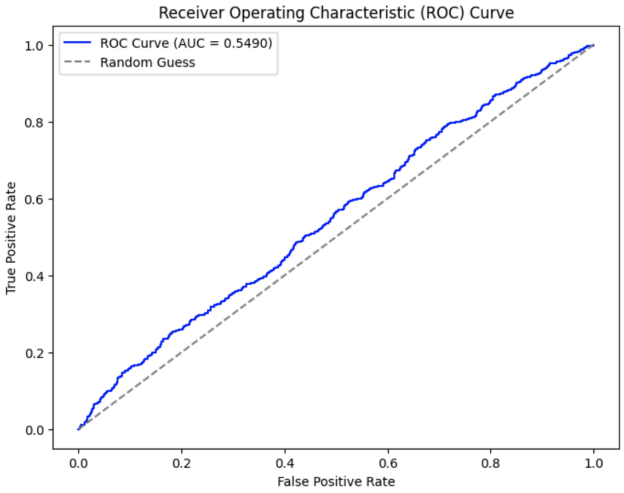
Libraries used:

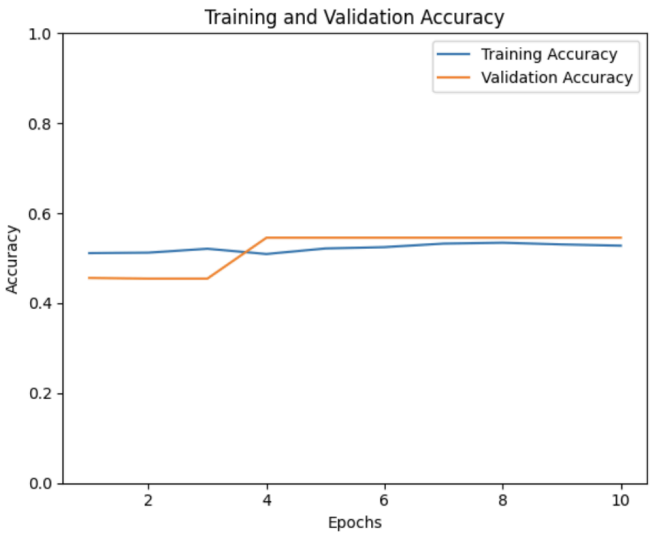
numpy - 1.12.6 | pandas - 2.2.3 | tensorflow - 2.16.2 | keras - 3.6.0 | opencv-python - 4.10.0.84 | matplotlib - 3.9.2 | imbalanced-learn - 0.12.4 | scikit-learn - 1.5.2 | seaborn - 0.13.2 | pillow - 11.0.0 | imagehash - 4.3.1

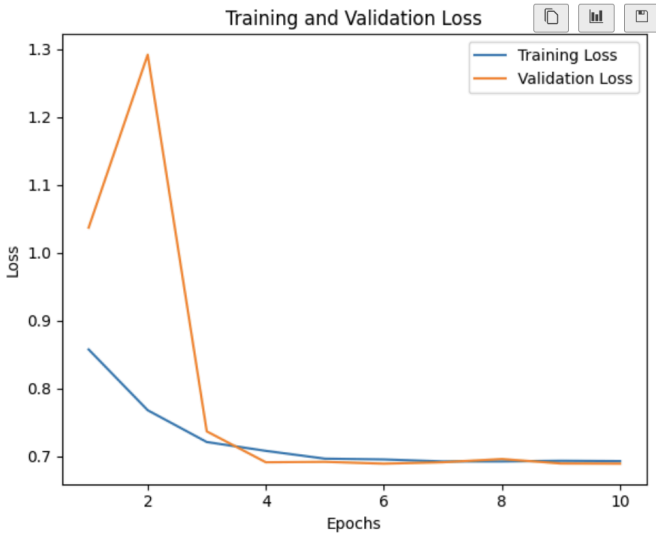
Additional Graphics

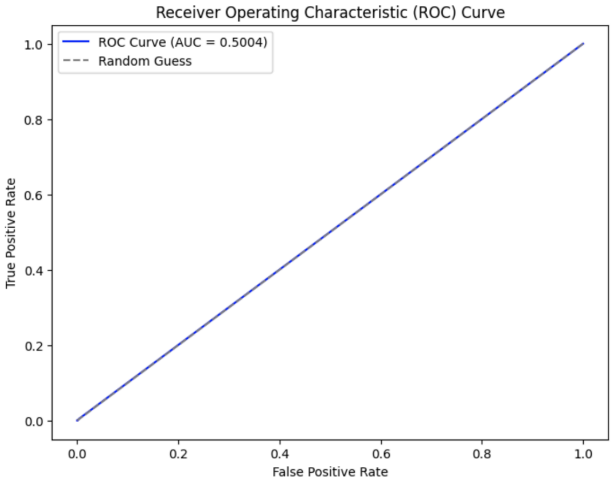


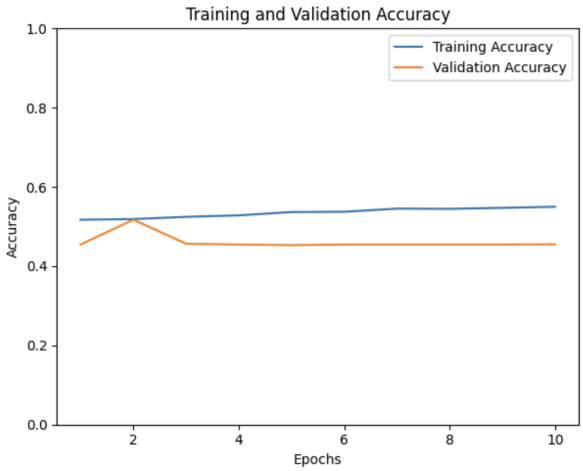


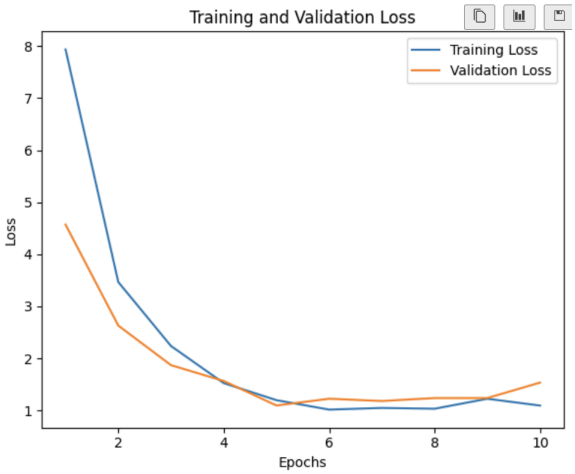


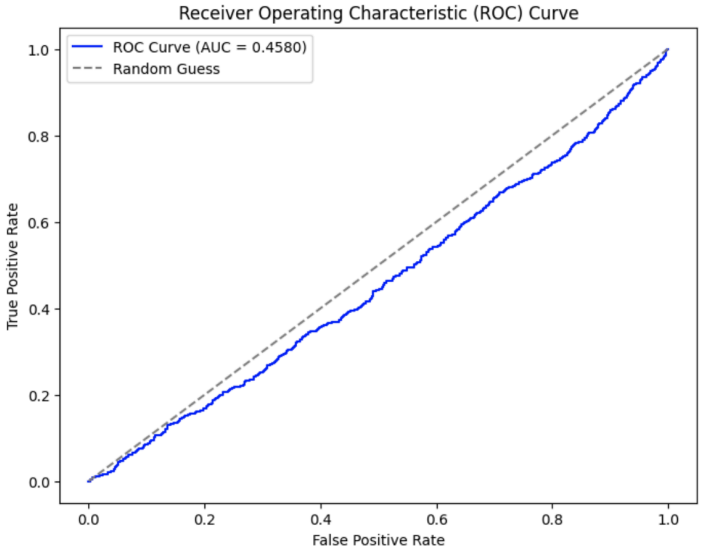


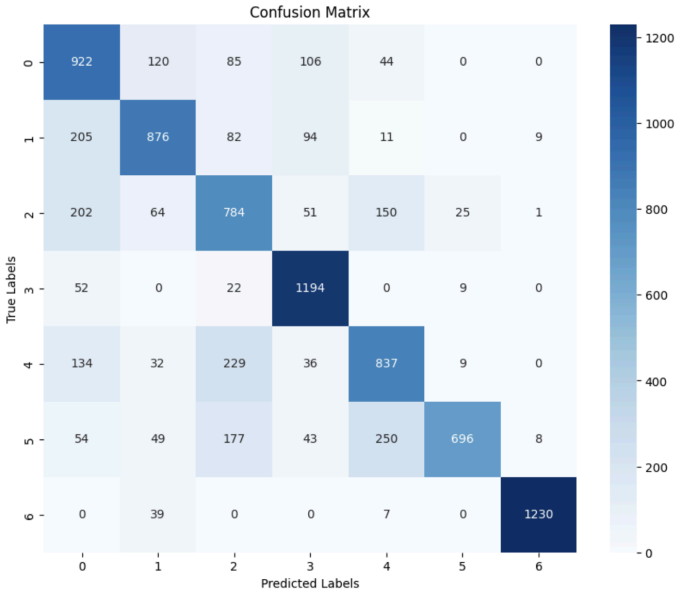



Receiver Operating Characteristic (ROC) Curve
ROC Curve (AUC = 0.5004)
Random Guess
True Positive Rate
False Positive Rate

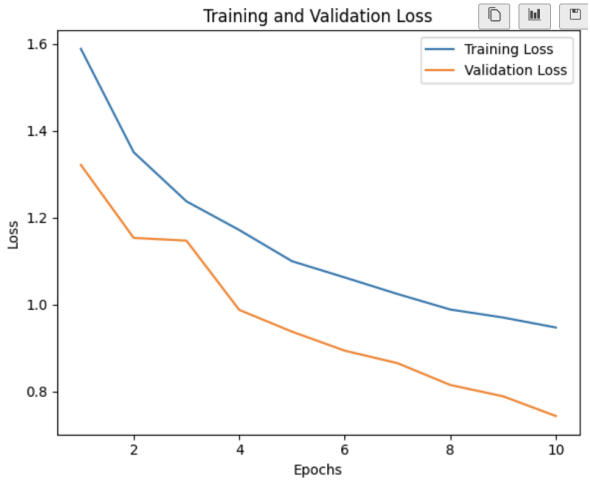

Training and Validation Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs

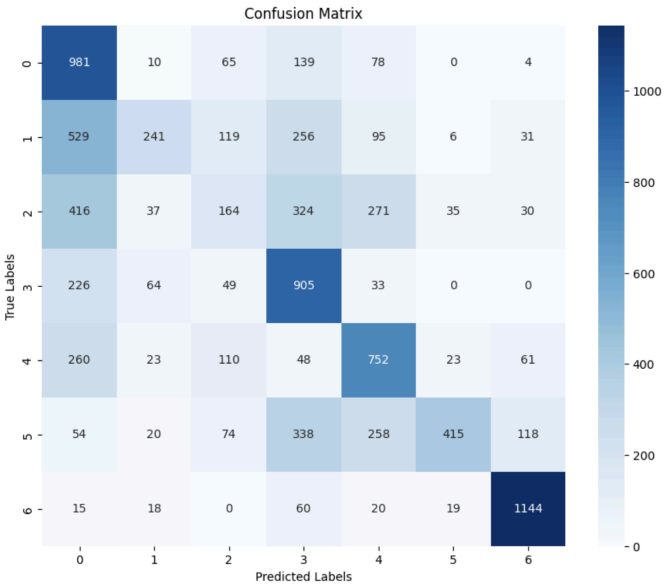

Training and Validation Loss
Training Loss
Validation Loss
Loss
Epochs

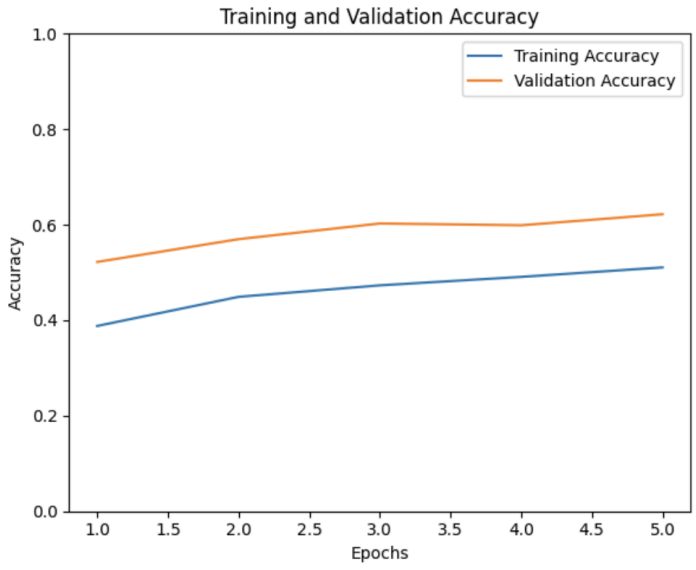

Receiver Operating Characteristic (ROC) Curve
ROC Curve (AUC = 0.4580)
Random Guess
True Positive Rate
False Positive Rate

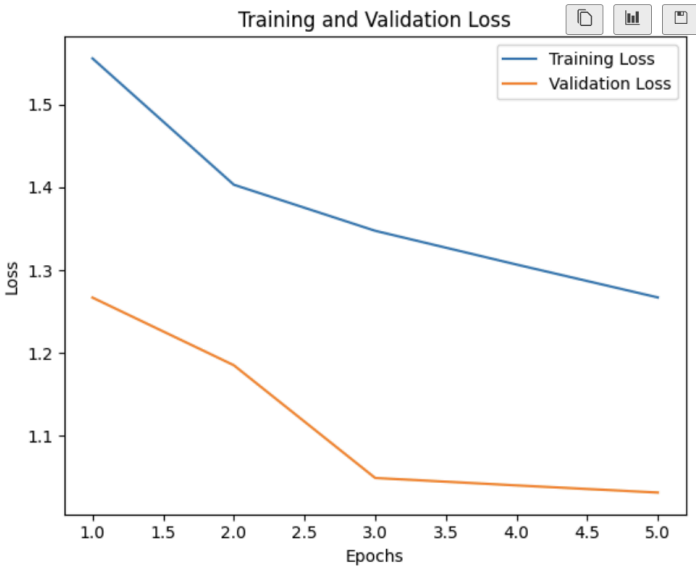

Confusion Matrix
True Labels
Predicted Labels
922 120 85 106 44 0 0
205 876 82 94 11 0 9
202 64 784 51 150 25 1
52 0 22 1194 0 9 0
134 32 229 36 837 9 0
54 49 177 43 250 696 8
0 39 0 0 7 0 1230

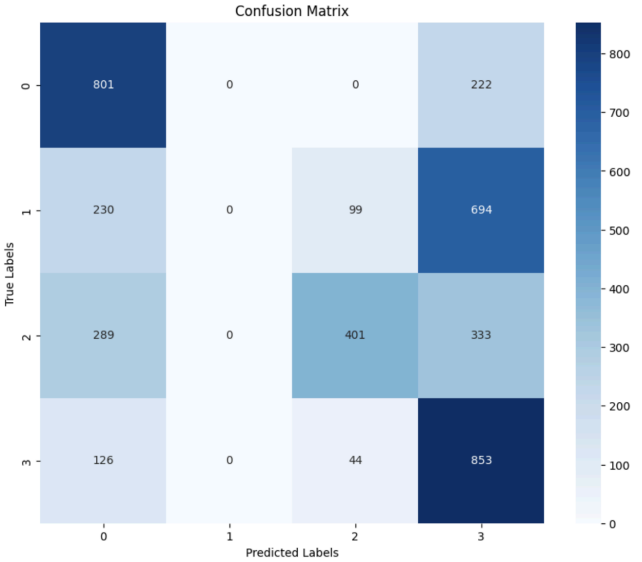

Training and Validation Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs

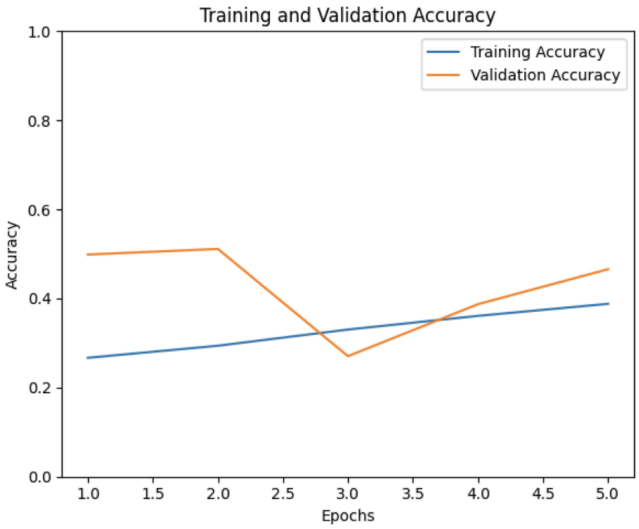

Training and Validation Loss
Training Loss
Validation Loss
Loss
Epochs

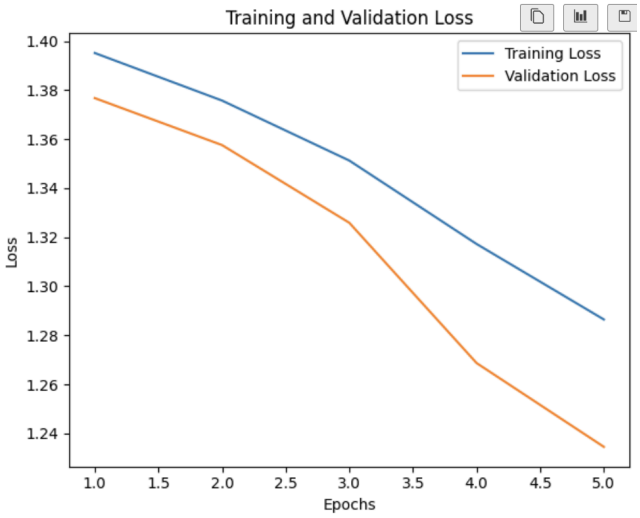

Confusion Matrix
True Labels
Predicted Labels
981 10 65 139 78 0 4
529 241 119 256 95 6 31
416 37 164 324 271 35 30
226 64 49 905 33 0 0
260 23 110 48 752 23 61
54 20 74 338 258 415 118
15 18 0 60 20 19 1144

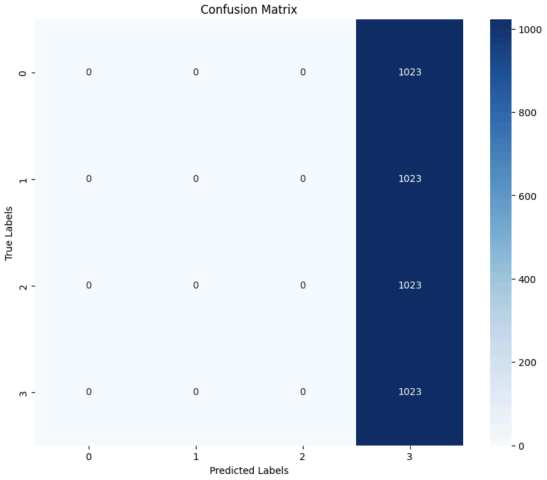

Training and Validation Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs

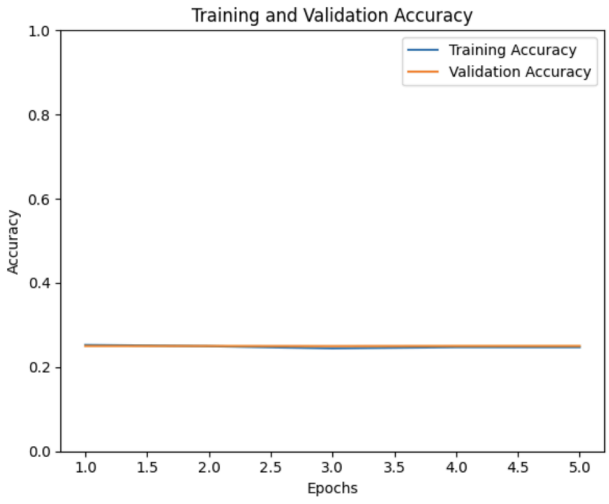

Training and Validation Loss
Training Loss
Validation Loss
Loss
Epochs

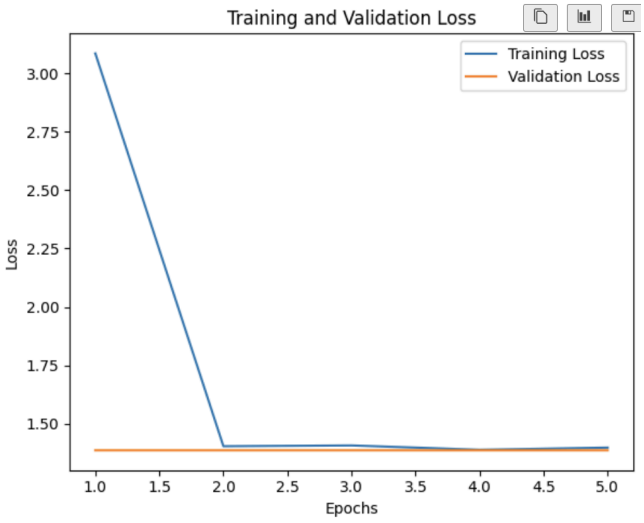

Confusion Matrix
801
0
0
222
230
0
99
694
289
0
401
333
126
0
44
853
True Labels
Predicted Labels

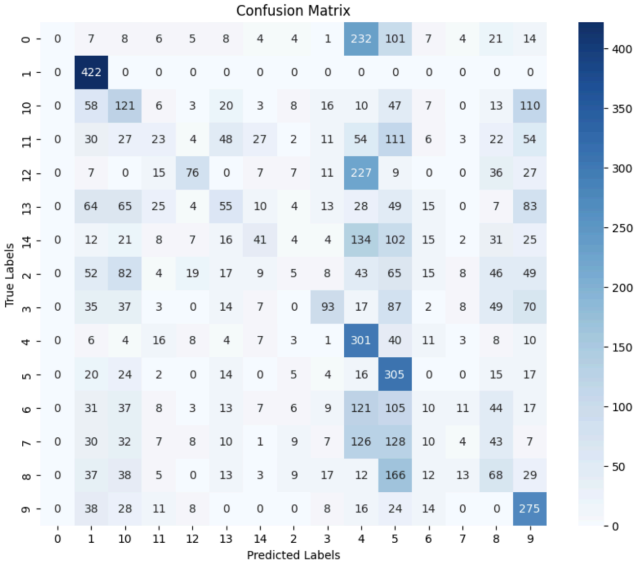

Training and Validation Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs

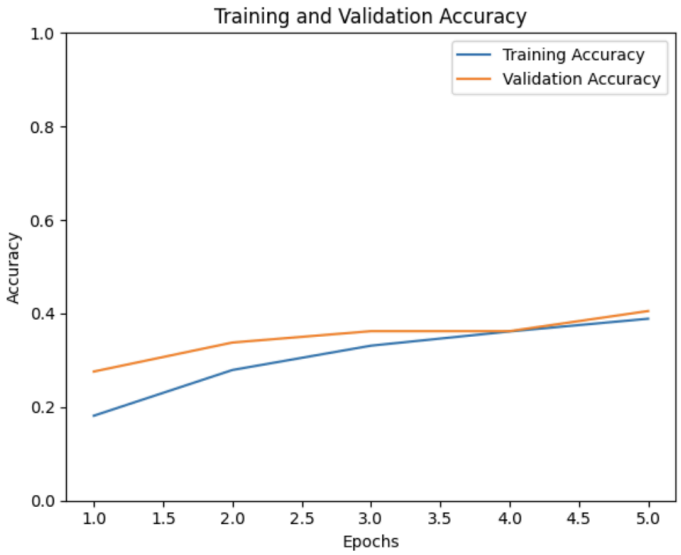

Training and Validation Loss
Training Loss
Validation Loss
Loss
Epochs

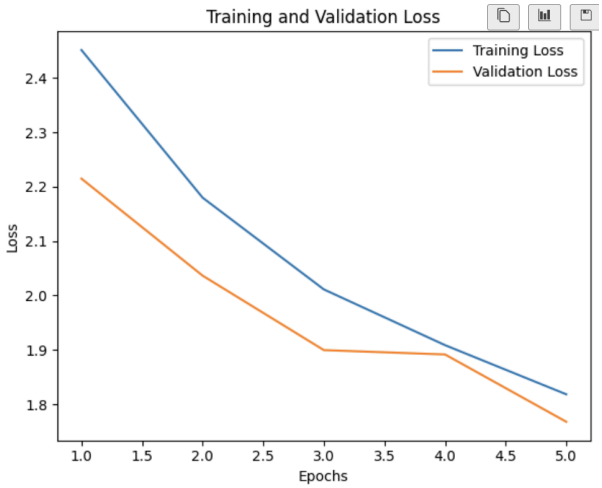

Confusion Matrix
1023
True Labels
Predicted Labels

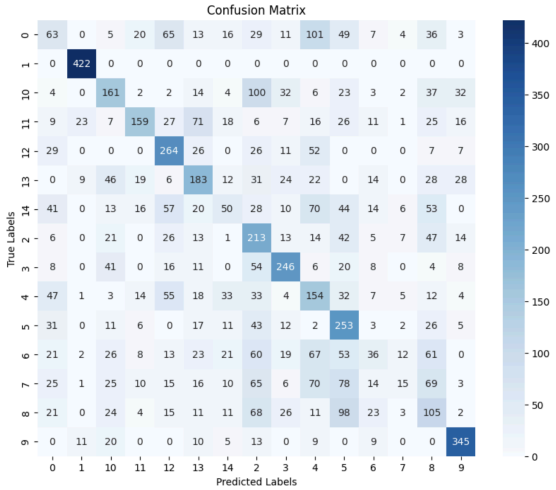

Training and Validation Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs

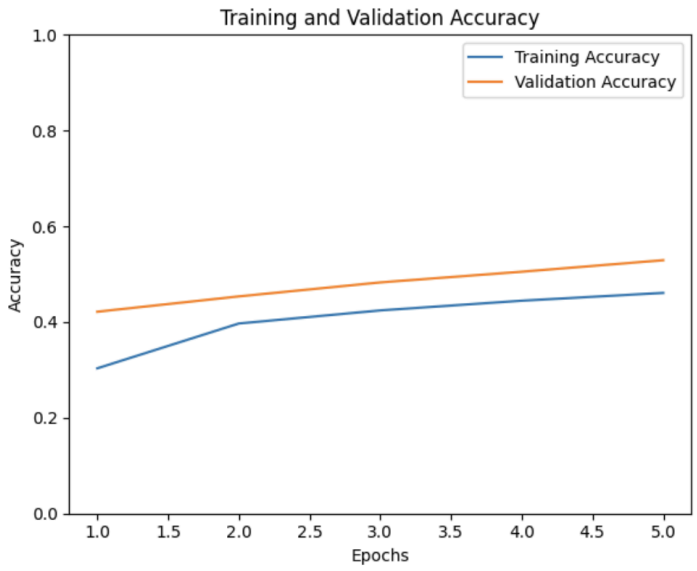

Training and Validation Loss
Training Loss
Validation Loss
Loss
Epochs

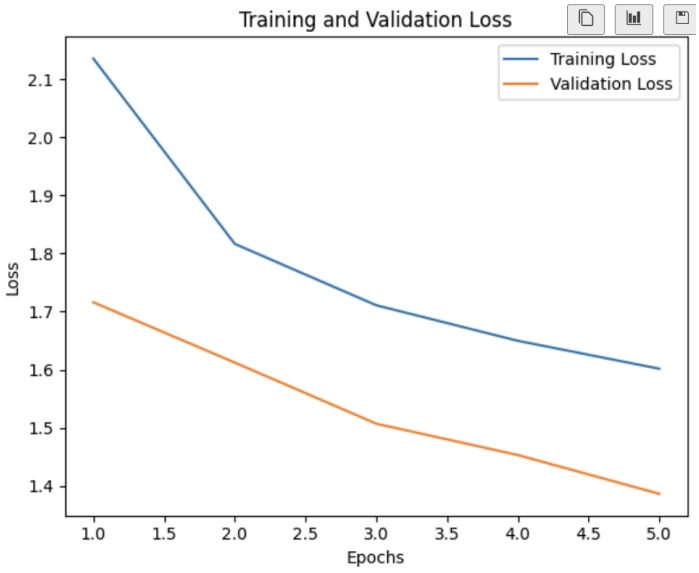

Confusion Matrix
True Labels
Predicted Labels


Training and Validation Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs


Training and Validation Loss
Training Loss
Validation Loss
Loss
Epochs


Confusion Matrix
True Labels
Predicted Labels


Training and Validation Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epochs


Training and Validation Loss
Training Loss
Validation Loss
Loss
Epochs