Deep Learning vs. Traditional Machine Learning for Real and Fake Image Classification

Nithin Reddy Nagapurr[1]

***Abstract*— This study compares the performance of a deep learning classifier (Convolutional Neural Network) and traditional machine learning models (Logistic Regression, SVM, Decision Tree, and Random Forest) in binary image classification for real and fake images. The results indicate that the deep learning classifier underperforms compared to traditional models, suggesting a need for improvements in its architecture or training methodology.**

I. INTRODUCTION

Binary image classification has become a crucial task in various applications, such as detecting real and fake images for security purposes or content moderation. Choosing the right model for the task is vital for achieving high accuracy and effectiveness. In this study, we aim to compare the performance of a deep learning classifier, specifically a Convolutional Neural Network (CNN), with traditional machine learning models like Logistic Regression, SVM, Decision Tree, and Random Forest.

II. DATASET AND PREPROCESSING

The dataset used in this study consists of real and fake images. We loaded the training and testing datasets from their respective directories and created a function to load a specified number of images from a directory and assign them to the given label. The loaded images were converted to the RGB format, and the function returned two arrays, one containing the images and the other containing their labels.

After loading the images, we combined them to form the final training and testing sets. To ensure a random distribution of real and fake images, the data was shuffled.

III. MODEL DEVELOPMENT

We developed a CNN model for the deep learning classifier, which consisted of three convolutional layers, two dense layers, and a final output layer for image classification. The model was compiled using the Adam optimizer and binary cross-entropy loss function.

In addition to the CNN model, we also implemented the following traditional machine learning models: Logistic Regression, SVM , Decision Tree, Random Forest.

IV. MODEL TRAINING AND EVALUATION

The CNN model was trained on the training set for 100 epochs. We saved the best model based on the validation loss for further evaluation. The model's performance was evaluated on the test set using the following metrics: accuracy, precision, recall, and F1 score.

We compared the performance of the CNN model with that of the traditional machine learning models.

V. RESULTS AND DISCUSSION

The CNN model's performance was significantly worse than that of the traditional machine learning classification models, with an accuracy of only 50%, precision of 25%, and recall of 50%. This suggests that the CNN model did not effectively learn to distinguish between real and fake images in the dataset.

Model	Accuracy	Precision	Recall	F1 Score
CNN (Deep Learning)	50%	25%	50%	33.3%
Logistic Regression	74%	75%	84%	76%
SVM	72%	73%	71%	72%
Random Forest Classifier	67%	67%	67%	67%

Fig. 1. Comparison of the CNN with traditional ML models

In comparison, all three traditional machine learning classification models performed significantly better than the CNN model, with accuracy ranging from 67% to 74%, precision ranging from 67% to 75%, recall ranging from 67% to 84%, and F1 score ranging from 67% to 76%. Logistic Regression performed the best overall, with the highest accuracy, precision, and F1 score, followed by the SVM and Decision Tree. The Random Forest classifier performed the worst among the traditional machine learning models.

It is important to note that the performance of each model can vary depending on the specific dataset and problem at hand. Therefore, it is crucial to evaluate and compare the performance of multiple models before making a final decision on which one to use.

VI. CONCLUSIONS

In conclusion, the deep learning classifier model did not perform well on the binary image classification task of distinguishing between real and fake images. Further improvement to the model's architecture or training methodology may be needed to enhance its performance. This study highlights the importance of evaluating and comparing multiple models, including both deep learning and traditional machine learning approaches, to select the most appropriate model for the task at hand. Future work may involve

[1]Nithin Reddy is a student of computer science, Wichita State University, Wichita, KS, 67214 nagapur@icloud.com