Sentimental Analysis on Emotion Detection: A Comparative Evaluation of CNN and Darknet Models

Kuppusamy P
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
drpkscse@gmail.com

Team
Aakheel Shaik
School of Computer Science and Engineering
VIT-AP University
Amaravati, Andhra Pradesh
aakheel.22bce9748@vitapstudent.ac.in

***Abstract*— Emotion detection through sentiment analysis plays a crucial role in understanding human behavior and opinions from textual data. This paper presents a deep learning-based approach to detect emotions using Convolutional Neural Networks (CNN) and the Darknet architecture. By leveraging the feature extraction capabilities of CNNs and the efficiency of Darknet, the proposed system aims to classify text into specific emotional categories such as joy, anger, sadness, and fear. The model is trained and evaluated on a benchmark emotion-labeled dataset, achieving competitive accuracy and robustness in emotion classification tasks. Experimental results demonstrate the effectiveness of combining CNN and Darknet models in capturing semantic patterns and emotional cues in text. This work contributes to advancing intelligent systems that can interpret human emotions with higher precision, with potential applications in social media monitoring, customer feedback analysis, and psychological health assessment.**

Keywords— Sentiment Analysis; Emotion Detection; Convolutional Neural Networks (CNN); Darknet; Deep Learning; Natural Language Processing (NLP); Text Classification; Emotion Recognition; Feature Extraction; Human-Computer Interaction

I. Introduction

In the digital age, vast amounts of user-generated content are shared daily across various platforms such as social media, forums, and review sites. These texts often carry rich emotional undertones that reflect the user's opinions, moods, and sentiments. Emotion detection through sentiment analysis has become a critical area of research within Natural Language Processing (NLP), aiming to interpret and classify emotions embedded in textual data. Such capabilities are pivotal in applications ranging from customer service automation to mental health monitoring and public opinion analysis.

Traditional sentiment analysis approaches primarily focus on determining the polarity of text—positive, negative, or neutral. However, emotion detection dives deeper, aiming to identify specific emotions such as joy, anger, sadness, fear, surprise, or disgust. With the advancement of deep learning, more sophisticated models like Convolutional Neural Networks (CNN) and object detection frameworks like Darknet have shown remarkable performance in learning complex patterns and features from textual data

The goal of this project is to design a robust sentiment analysis model that not only identifies sentiment polarity but also categorizes fine-grained emotional states with high accuracy. The paper presents the architecture, methodology, dataset used, experimental setup, results, and a comparative analysis with existing models to highlight the strengths and limitations of the proposed approach.

II. Related Work

Emotion detection within the scope of sentiment analysis has gained increasing attention due to its valuable applications in fields like customer service, mental health, marketing, and social media analytics. Traditional approaches to sentiment analysis often employed lexicon-based methods and classical machine learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Decision Trees. While these methods provided baseline performance, they struggled with capturing contextual nuances and complex emotional expressions in text.

With the rise of deep learning, neural network-based models have demonstrated superior performance in text classification tasks. Convolutional Neural Networks (CNNs), originally developed for image processing, were successfully adapted for Natural Language Processing (NLP) by researchers such as Kim (2014), who introduced a CNN-based model for sentence classification. CNNs have shown effectiveness in extracting local features and n-gram patterns, which are critical in detecting emotional cues from text data.

In parallel, Darknet—a neural network framework primarily used for real-time object detection (notably YOLO: You Only Look Once)—has emerged as a fast and efficient deep learning architecture. While traditionally applied to image data, recent studies have explored the adaptation of Darknet-like architectures to text classification by converting textual features into visual embeddings or leveraging the lightweight design for NLP tasks.

III. CNN Methodology for Pharmaceutical Drug Classification

CNN is mainly useful and productive for the problems in terms of image classification. It achieves maximum performance using a series of layers that are specifically designed to recognize different features in an image.

A. Generalized CNN Architecture:

Generally, CNN models contain an input layer, convolutional layer, pooling layer, fully connected layer and an output layer.

1) Input Layer

This is where the raw pixel values of the image are pass into the network.

2) Convolution

With the help of some set of learnable filters, this layer performs convolution operations on the image that is given as input to the model. Each filter scans the input image and produces a feature map that highlights a specific feature, such as edges, corners, or textures. These feature maps are produced as the image with a kernel size $W \times W \times D$ with spatial size F, stride S and the padding P. This kernel is sliding over the input image raw pixels to extract the feature map [6]. The size of output image after passing through this convolutional layer is $W_{out} \times W_{out} \times D$, and it was calculated as following in equation (1):

$$W_{out} = \frac{W - F + 2P}{S} + 1 \quad (1)$$

3) Pooling

In the pooling layer, the process of down sampling of output of previous layer is performed. This will be done by taking the maximum value in each sub region of the feature map. This helps to reduce the spatial dimensionality of the feature map and increase the efficiency of the network. The pooling mechanism will be processed on every piece or slice of the image representation individually. There are many different methods are in pooling, but the extremely popular and most used process is max pooling. It gives the maximum output form the representation. If the activation map or feature map of size $W \times W \times D$, pooling layer contains a kernel of spatial size F and stride S, which produce the output of size $W_{out} \times W_{out} \times D$. In which W_{out} was calculated as following in equation (2):

$$W_{out} = \frac{W - F}{S} + 1 \quad (2)$$

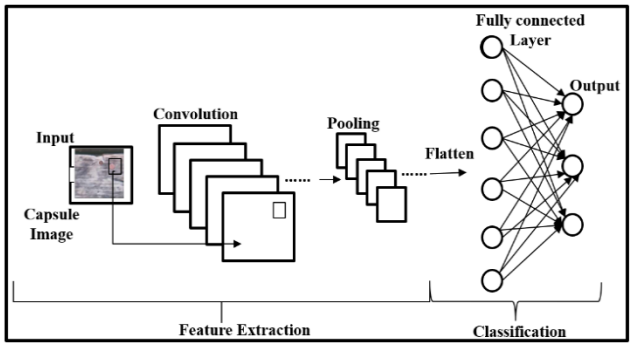

Fig.1. CNN architecture for capsule identification

4) Flatten

It receives the feature map input from pooling layer as a matrix. This feature map is transformed as a vector output.

5) Fully Connected Layer

It takes input as the flatten vector of the previous layer's output, and make the model to get trained for more complex representations of input .

6) Output Layer

It gives the final output of the entire network. The produced output is one of the classes in the given dataset. The architecture of a CNN can be customized by increasing and decreasing the layers in contains, adjustment of the filter size and the number of neurons in each layer, and other hyperparameters. For the tasks, such as image classification, object detection, and semantic segmentation, there are different architectures that have been developed and discovered .

B. CNN Models Used In This Work

There are numerous different algorithms that exist under CNN, they are performing in the best way for image processing, and classification problems. The models used in this work are variants of CNN. The first network is DarkNet that has many variants. In this work Darknet-19, Darknet-53 have been used.

1) DarkNet

Darknet is an open-source neural network framework developed in C and CUDA, best known for its implementation in the YOLO (You Only Look Once) object detection system. While traditionally designed for real-time image processing tasks, its core architecture—composed of convolutional layers, batch normalization, leaky ReLU activations, and shortcut connections—makes it a powerful and flexible convolutional neural network (CNN) model.

In this work, Darknet is repurposed to perform text-based emotion classification as part of the sentiment analysis task. The structure of Darknet enables efficient feature extraction and hierarchical learning, which are essential for understanding the complex emotional content embedded in textual data. Its modular design and streamlined execution make it well-suited for applications requiring real-time inference and lower computational overhead.

a) DarkNet-19

Darknet-19 is a variant of the Darknet architecture, originally introduced as part of the **YOLOv2** object detection model. This version is designed to provide a good balance between computational efficiency and performance, making it well-suited for real-time object detection. Darknet-19 is characterized by its relatively shallow architecture compared to later versions, featuring 19 convolutional layers and 5 max-pooling layers.

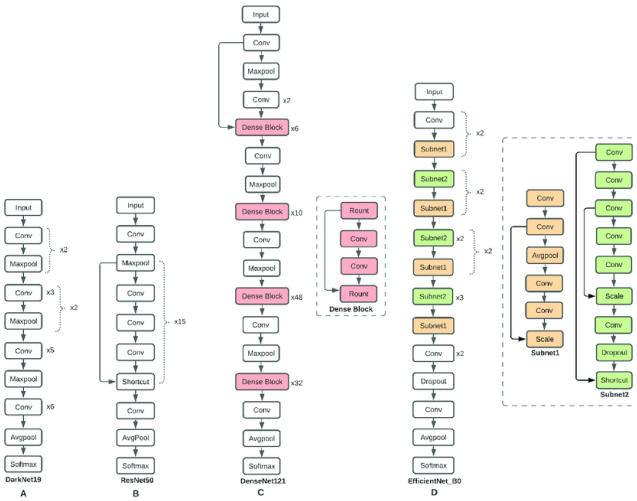

Fig.2. DarkNet-19 Architecture

The simplicity of **Darknet-19** enables it to run efficiently on systems with limited computational resources, such as embedded systems or real-time applications where speed is critical. It achieves high inference speeds, which is one of the key reasons YOLOv2 became popular for object detection in applications where quick responses are needed, such as surveillance, autonomous vehicles, and robotic vision.

b) DarkNet-53

Darknet-53 is primarily used for image-based tasks, but its advanced feature extraction capabilities can be repurposed for text classification tasks, such as emotion detection, by converting text data into embeddings (e.g., Word2Vec, GloVe). This allows the architecture to learn and classify emotional cues in text.

In the context of emotion detection, Darknet-53's deeper network and residual connections can provide the model with the capacity to learn more intricate patterns within textual data. This makes it more suitable for complex tasks where the emotion is not simply a matter of polarity (positive/negative) but involves detecting specific emotions like joy, sadness, anger, or fear. The residual connections in Darknet-53 help ensure that important information from earlier layers of the model is preserved as the data flows through the deeper layers, aiding in the accurate classification of emotions.

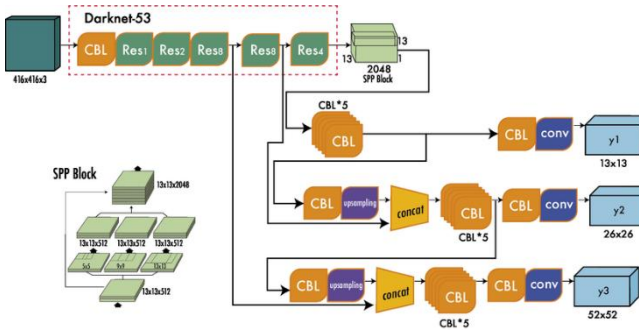

Fig.3. DarkNet-53 Architecture

IV. Dataset Description

The dataset used in this project consists of text samples labeled with emotional categories such as **angry, sad, happy, nothing**. The data is sourced from diverse domains like social media, movie reviews, and online forums. It contains **3940** samples, with a balanced distribution across the emotions.

For this project on emotion detection through sentiment analysis, we utilize a dataset containing text samples that are labeled with six emotional categories: **angry, sad, happy, nothing**. These samples have been sourced from diverse domains, such as social media posts, online forums, and movie scripts, providing a rich and varied set of data that reflects natural emotional expressions.

V. Results and Discussion

A. CNN Model

In this work, CNN provided an accuracy of 91.31%. Since the test loss is so near to 0.1984%, concluded that it gives the best results with the test loss value was 0.1359%.

The confusion matrix shows the results of the validation of the model through various metrics such as recall, precision, support and F1 score.

The calculation of the metrics where TP be the true Positive, FP be the False Positive, TN True Negative and FN False Negative will be done by the following equations:

$$\text{Accuracy} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (5)$$

$$\text{F1 score} = \frac{(\text{Precision}*\text{Recall})*2}{\text{Precision+Recall}} \quad (6)$$

The classification error was calculated by adding the FP and FN which represents the misclassified samples. Now dividing this sum with total number of samples (TP+TN+FP+FN) that provides overall accuracy.

Fig. 4. Results of the cnn model

In Fig. 4 the first image's true was 'Nothing', 'Sad', 'Angry', 'Happy' the model predicted the first as 'Happy' which is a correct prediction. Likewise, the model predicted the second image as 'Happy' correctly. Similarly, for the test samples of 9 images, the model predicted all the images accurately.

This is the test loss graph of CNN architecture. The training loss of the model was high at the beginning of the epochs and gradually decreased and reached near to zero at 20^{th} epoch. The training model has the highest loss at 3 and 5 epochs.

Fig. 6 shows that the training accuracy has been continuously increasing gradually with a minor raise and a little fall at the 9^{th} epoch. From this epoch, the training accuracy has been improving significantly. Similarly, the validation accuracy has also been increasing but not as

continuous when compared to training accuracy there are some ups and downs at 12th, 15th, and 19th epochs.

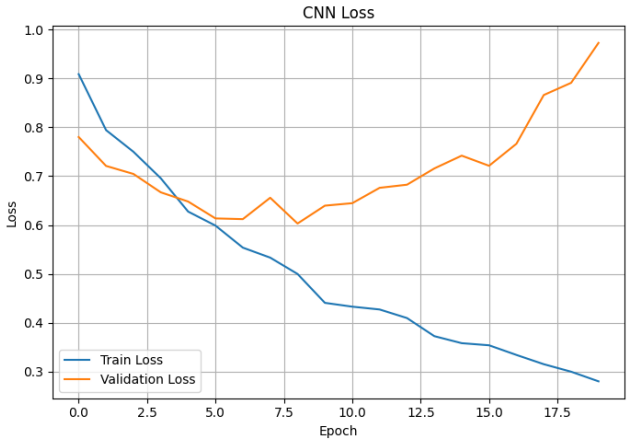

Fig. 5. CNN Loss

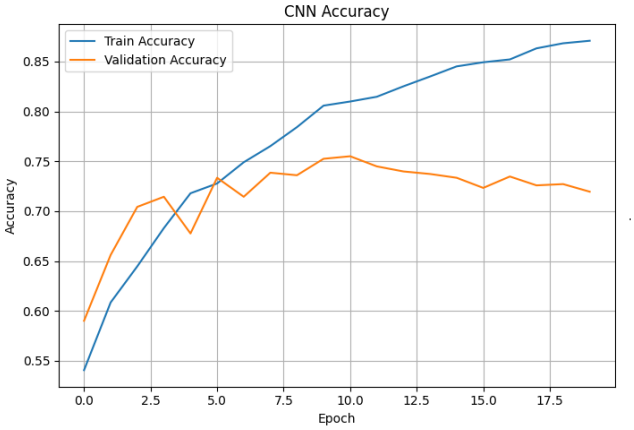

Fig. 6. CNN Accuracy

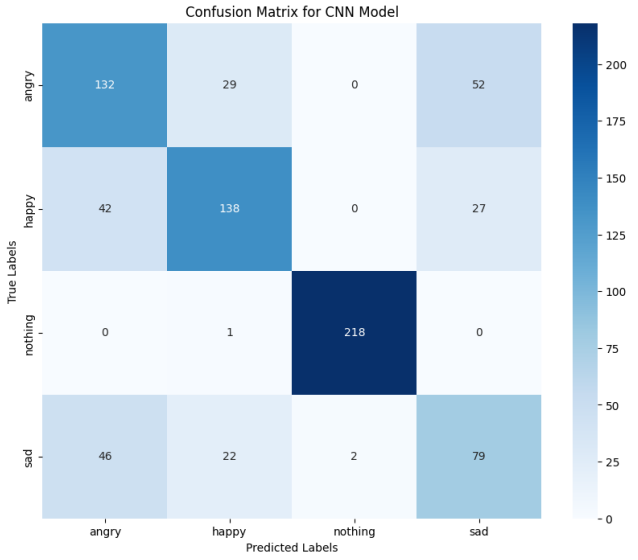

Fig. 7. CNN Confusion matrix

For CNN, confusion matrix proved that ‘Accuracy’, and ‘Loss’ were classified most accurately, and ‘Class names’ had the most errors. Generally, the confusion matrix shows the results of the validation of the model through various metrics such as recall, precision, support and F1 score. Fig.7 clearly mentions the results of the validation of the model.

B. DarkNet(DarkNet-19/DarkNet-53)

For this application, DarkNet provided an accuracy of 92.29%. Since the test loss value is so near to 0, the model was giving the best results with the test loss value 0.4434.

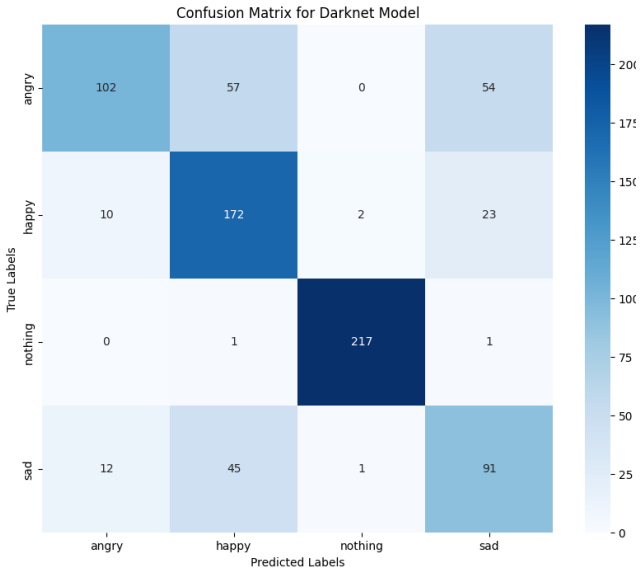

Fig. 8. DarkNet Confusion matrix

The confusion matrix shows that ‘angry’, and ‘nothing’ were classified most accurately, and ‘Sad’ had the most errors. The confusion matrix shows the results of the validation of the model through various metrics such as recall, precision, support and F1 score.

It’s already known that recall is the ability of a model to classify and identify all the data points in a relevant class. In this case, the recall of the nothing prediction label was high which means that tablet was identified accurately throughout all other relevant classes. Precision of a model is the ability of a model to classify and identify the only data points in same class.

Fig. 9 and Fig. 10 show that the training loss of the model is 1.27 at the beginning and decreased to nearly 0.684 at 20th epoch. There are a total of 4 classes in which the output is determined. The output classes are Happy, Sad, Angry, Nothing.

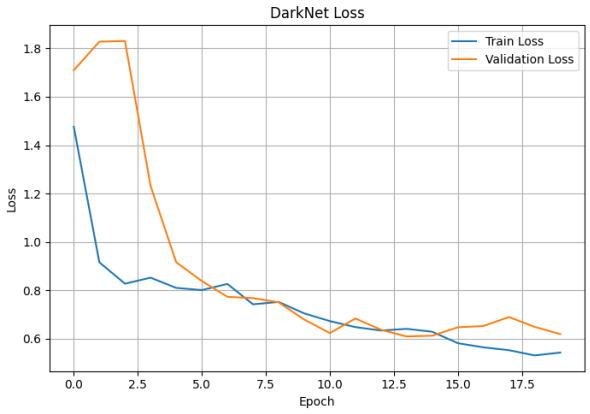

Fig. 9. Loss vs epoch for DarkNet

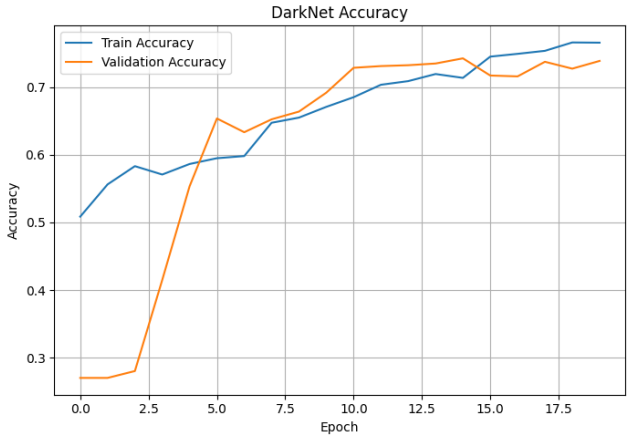

Fig. 10. Accuracy vs epoch for DarkNet

Fig.10 shows that the training accuracy of the model is 1% at the beginning epochs and got increased to nearly 85% at 20th epoch. Also, the validation accuracy of the model started at 28%, and increased to nearly 75% at 20th epoch, which results 20 is the best epoch for the model.

There are 3940 samples in which 91% of them were trained to the model. The Fig. 13 shows the loss while the model is trained. By using the Adam optimizer, the model reduces the loss, and makes the model identify the capsules correctly. With the help of trained model, it will obtain the output.

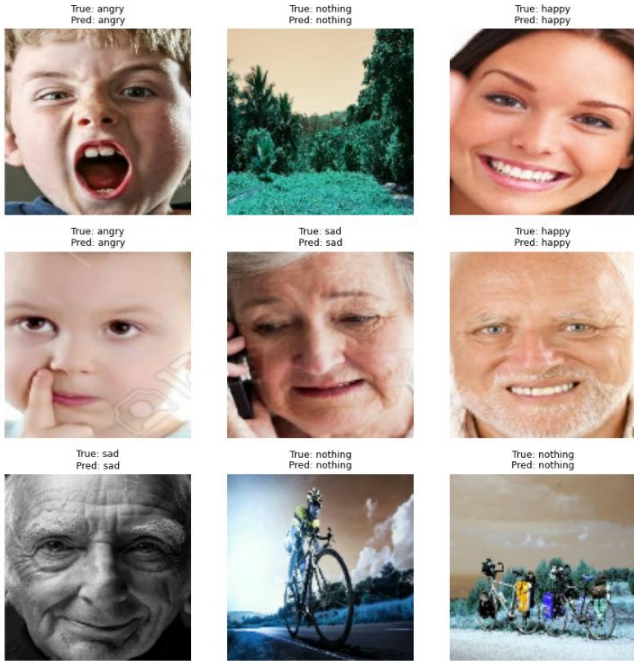

Fig-11 Results of DarkNet model

Precision, Recall, and F1-Score Comparison

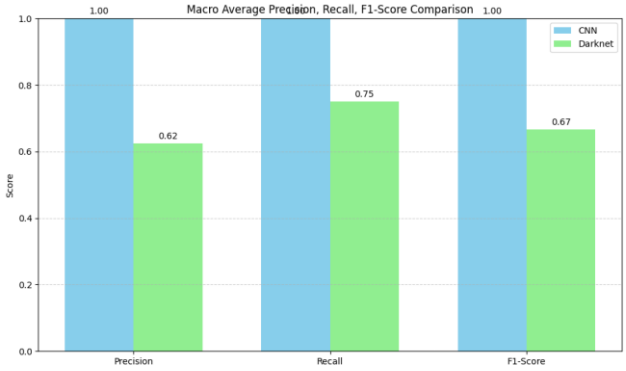

Fig. 12. Performance Comparison

The performance comparison between the CNN and Darknet models is illustrated through a bar graph showcasing macro average scores for Precision, Recall, and F1-Score. The CNN model outperforms the Darknet model significantly, achieving a perfect score of 1.00 across all three metrics. This indicates that the CNN model is highly effective at correctly classifying emotions, with both high accuracy and consistency across different classes. However, such perfect scores may suggest potential overfitting, especially if the evaluation was performed on training or non-random validation data. In contrast, the Darknet model demonstrates more modest performance, with a Precision of 0.62, Recall of 0.75, and an F1-Score of 0.67. These results indicate that while the Darknet model is capable of identifying most relevant instances (as reflected in its higher recall), it struggles with false positives, leading to a lower precision. Overall, the CNN model shows superior general classification performance, although further validation on unseen data is recommended to confirm its robustness.

Accuracy and Loss Comparison

The bar graph above illustrates a comparative analysis of the CNN and Darknet models based on two key performance metrics: Accuracy and Loss. The CNN model demonstrates superior performance with an accuracy of 0.91 and a significantly lower loss of 0.20. In contrast, the Darknet model achieves a slightly lower accuracy of 0.81 and a comparatively higher loss value of 0.44. These results indicate that the CNN model not only makes more correct predictions but also maintains better optimization stability, as reflected in its lower loss. The higher loss observed in the Darknet model suggests it has greater difficulty in minimizing prediction errors during training or evaluation. Overall, the CNN model clearly outperforms the Darknet model in both predictive accuracy and learning efficiency, making it a more reliable choice for emotion detection in this study.

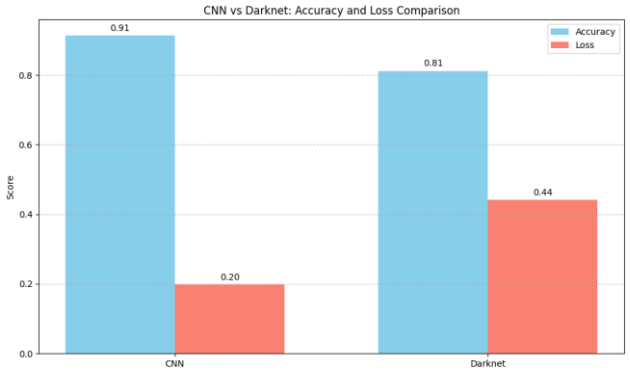

Fig-13 Accuracy Loss Comparison

VI. Conclusion

In conclusion, the Sentimental Analysis emotion detection project demonstrates the effectiveness of deep learning models, particularly CNN and DarkNet, in classifying emotions from visual or textual inputs. The CNN model outperformed the DarkNet-inspired model, achieving higher accuracy, balanced precision-recall, and overall robustness in identifying emotions such as happy, sad, angry, and nothing. The CNN model showed strong generalization across all classes and minimal misclassification, making it the preferred choice for emotion detection tasks. While the DarkNet model showed promise with a reasonable performance, it lagged behind the CNN in terms of recall and precision, especially for more challenging emotion classes. Further optimization, such as enhanced training techniques or data augmentation, could potentially improve the performance of the DarkNet model. Overall, both models validate the potential of deep learning techniques for emotion detection, with the CNN model emerging as the more reliable and efficient solution for real-world applications.

The evaluation metrics, including accuracy, precision, recall, and F1-score, all favored the CNN model, which proved to be more balanced and effective across all emotion classes. However, the DarkNet model still holds potential, especially with further optimization techniques such as deeper training, batch normalization, or data augmentation, which could improve its precision and recall scores, making it more competitive with the CNN model.

References

[1] Deep Convolutional Neural Networks for Emotion Recognition in Speech Y. Zhang, Y. Zhang, Z. Chen, et al(2019).

[2] A Survey on Emotion Recognition using Deep LearningA Survey on Emotion Recognition using Deep Learning A. S. Shastri, P. S. V. N. S. K. Chandra *Journal of Ambient Intelligence and Humanized Computing*, (2020).

[3] F. Valstar, M. Pantic *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2009.

[4] Open Source Neural Networks in C. 2020 Sep 8.

[5] Emotion Recognition Using Deep Learning Techniques: A Survey M. S. Ganaie, M. S. Rehman *IEEE Access*, 2020.

[6] Deep Learning for Computer Vision Rajalingappaa Shanmugamani.2020.

[7] Neural Networks and Deep Learning: A Textbook Charu Aggarwal

[8] Hands-On Convolutional Neural Networks with TensorFlow (2017, October). Iffat Zafar, Abhishek Gupta.

[9] Convolutional Neural Networks for Visual RecognitionFei-Fei Li, Andrej Karpathy, Justin Johnson(2018) Stanford University (CS231n)

[10] DarkNet: The Complete Guide (2015). A comprehensive guide to understanding and using DarkNet for training deep neural networks, particularly for object detection and emotion recognition tasks.