# Understanding Art Deeply: Sentiment Analysis of Facial Expressions of Graphic Arts Using Deep Learning

Fei Wang[1]
Hubei University of Technology, Wuhan City, Hubei Province, 430068, China[1]

***Abstract*—Art serves as a profound medium for humans to express and present their thoughts, emotions, and experiences in aesthetically and captivating means. It is like a universal language transcending the limitations of language enabling communication of complex ideas and feelings. Artificial Intelligence (AI) based data analytics are being applied for research domains such as sentiment analysis in which usually text data is analyzed for opinion mining. In this research study, we take art work and apply deep learning (DL) algorithms to classify seven diverse facial expressions in graphics art. For empirical analysis, state of the art deep learning algorithms of Inceptionv3 and pre-trained model of ResNet have been applied on large dataset. Both models are considered revolutionary deep learning architecture allowing for the training of much deeper networks and thus enhancing model performance in various computer vision tasks such as image recognition and classification tasks. The comprehensive results analysis reveals that the proposed methods of ResNet and Inceptionv3 have achieved accuracy as high as 98% and 99% respectively as compared to existing approaches in the relevant field. This research contributes to the fields of sentiment analysis, computational visual art, and human-computer interaction by addressing the detection of seven diverse facial expressions in graphic art. Our approach enables enhanced understanding of user sentiments, offering significant implications for improving user engagement, emotional intelligence in AI-driven systems, and personalized experiences in digital platforms. This study bridges the gap between visual aesthetics and sentiment detection, providing novel insights into how graphic art influences and reflects human emotions by highlighting the efficacy of DL frameworks for real-time emotion detection applications in diverse fields such as human psychological assessment and behavior analysis.**

***Keywords—Artificial intelligence; deep learning; sentiment analysis; art detection; image processing; convolutional network***

## I. Introduction

Art is a broad term that can be described as the work or process undertaken by man in creating physical skills, objects, or musicals involving painting, sculpture and dancing etc. They are not only appraisals of culture and individual encounters but also as a channel or medium of communication which enforces feeling and thinking. Interdisciplinary methods are used in the analysis of art that many disciplines incorporate into their analysis the impact and role of artistic creations on and from the social contexts of world. For example, in expressions analysis of art, the concern is on the feelings invoked by art and how such feelings differ among the subgroups and cultures [1]. In more practical terms, researchers use semi structured interviews and questionnaires with the audience together with quantitative tools like sentiment analysis to elicit responses that contribute to a nuanced understanding of the use of art as a means of creating human experiences and engagement with various social processes. There are various types of sentiment analysis including the binary classification of subjectivity analysis [2], the tertiary classification containing the sentiment valance finding [3], the multi-classification containing the emotion detection [4], and the aspect-oriented sentiment analysis [5] that targets feature level deep understanding. a means of communication that evokes emotions and provokes thought. The analysis of art spans multiple research areas, including psychology, sociology, and cultural studies, where scholars examine how artistic expressions influence and are influenced by societal contexts. For instance, expressions analysis in art focuses on understanding the emotional responses elicited by artworks and how these responses vary across different demographics and cultural backgrounds [6]. Researchers employ qualitative methods such as interviews and surveys alongside quantitative techniques like sentiment analysis to gauge audience reactions, ultimately contributing to a deeper understanding of the role of art in shaping human experience and social discourse [7].

Moreover, sentiment analysis is combined with other technologies like computer vision and IoT devices so that a business can monitor information about customers’ emotions and actions in the physical spaces in real-time mode [8]. Sentiment analysis is an important area as it continues to develop, the complexity of the models for such analysis will increase making it easier to determine the right strategies when they are required in different fields. Analysis of sentiment in images has been a popular trend in recent years mainly in the context of affective content in images. This field discusses how certain images create certain feelings and this is very essential because in areas like social media analysis or advertising. The research in this direction started around 2010, where the first attempts were made to place pictures into positive or negative sets according to their characteristics. To interpret affect, there has been the use of methods like Convolutional Neural Networks (CNNs) to perform context analysis of images, in relation to emotional content through organizations like Flickr and Twitter. From the research done, texture and color co-occurrence histogram are critical in identifying the sentiment of an image [9] [10]. In addition, proposing a method for combining both text and image features should help improve the effectiveness of sentiment classifiers and provide additional knowledge of the users' emotions conveyed through images posted on social media [11].

As for social art, sentiment analysis becomes a crucial method on how the public perceives the art pieces and other materials that are a part of culture. By observing images of artworks or social art initiatives posted on social media, emotions and reception of a given subject in the community can be quantified [12]. This approach can help in measuring the level of audience participation but can also enlighten artists and curators regarding prevailing mood trends within their viewers. When applied in this case, the deep learning models enable the analysis of subtle differences in sentiment beyond basic positive-negative quality assessments [13]. In addition, the differentiation of emotions that are related to concrete artworks will enable targeted addressing and, thus, improve the effectiveness of social art activities.

### A. *Research Contributions*

In this study, our main contributions include:

- For graphic art and identification of seven emotions, data preprocessing and diverse deep learning algorithms have been applied.
- Highlighted the limitations of existing studies by filling research gaps in emotion detection using digital image processing and deep learning.
- Developed a robust emotion classification framework using a deep learning pre-trained models including ResNet-50 and Inception v3 model by modifying the fully connected layer to adapt to the specific emotion classification task.
- Achieved highest classification performance of 99% with inception v3 model as compared to baseline models such as VGG16 and DCNN, demonstrating state-of-the-art results.

For the rest of the paper, Section II reviews the existing studies in relevant literature, then Section III shares the proposed research methodology along with experimental set-up discussing datasets which are prepared and used for empirical analysis and performance metrics used for results comparison. Section IV presents the results and discussion sharing findings from this study. Before concluding the manuscript, Section V discusses the results in detail sharing comparative analysis of the proposed model with the existing approaches.

## II. Background

The sentiment analysis of images by employing deep learning techniques has received increased attention in the last few years due mainly to the large availability of computer powers and the growing availability of image data in the social media platforms. In this approach, deep learning techniques, including Convolutional Neural Networks CNNs, are used to learn useful feature representations of images from social media which are indicative of emotional sentiments. Table I defines the summary of existing studies for deeper analysis. Studies show that CNNs are capable to capture spatial hierarchies of images for the task of categorizing sentiments into positive, negative or neutral [14]. Recent works have shown that using transfer learning methods including Inception-V3 it is possible to librarian the accurate classification of the sentiment analysis tasks without requiring large, labeled datasets [15]. This capability is particularly valuable where good quality labeled data is hard to come by or in short supply.

The combination of two modalities, i.e., using image analysis in conjunction with textual sentiment analysis, has enriched the field even more. Analyzing the pictures together with the related texts helps researchers get a deeper insight into the people's attitudes [16]. For example, the integration with captions or hashtags used in Big Five Personality traits analysis helps models consider extra context that enhances the sentiment prediction accuracy up to multiple factors [17]. Moreover, improvements in the deeper architecture of the Capsule Networks besides the convolutions with RNN and hybrid Deep Learning also demonstrate great performance in sentiment analysis [18]. These models prove enhanced performances in comprehending intricate nonverbal emotions described through graphics. There is still a problem in image sentiment analysis, especially in terms of the stability of human emotions and different perceptions of images across cultures. Due to its subjectivity, there are variations in modeling sentiments which need to be dealt with by having rich training set with various emotions and occurrence [19]. A revolution in the recent few years in deep learning has revolutionized the field of image sentiment analysis where the general expressions of emotions in the image are processed and understood [20].

Among the more significant trends, it is possible to distinguish the combination of CNNs and RNNs as these networks have been used to extract spatial and temporal data in images and their textual descriptions. CNNs are good at detailing the local features protruding on architectural diagrams that make them significant in accentuated sentiment classifying assignments where visualization features are dominant [21]. Current research has also shown that combining CNNs with LSTMs results in finer outcomes in the sentiment classification owing to combining pros of both structures [22].

The use of people's emotions in multimodal textual and visual platforms has been considered as an active research area in this context. Combining information retrieved from both images and associated captions will give more light to researchers to get to the core point of this subject, which in this case is the sentiments. For example, it is established that the interaction of CNNs for image processing with individualized LSTMs or transformers for sports sentiment analysis can help improve sentiment outlook than individual modality alone [23]. This approach is especially useful in settings such as social media where messages are occasionally accompanied by images which indicate the authors' emotional state. Thus, the current state of and future trends for image sentiment analysis based on deep learning methods are steadily developing. We find a vast scope towards improving the existing sentiment analysis performance and its utility in different domains by integrating them with the latest architectural models like CNNs, RNNs, and transformers with a combination of multi-modal data for the prediction of gender violence based on sentiment analysis [24]. As the issues concerning data quality and ethical implications are solved as potential issues, it will be possible that both quantity and quality aspects of deep learning-based sentiment analysis will expand.

### *A. Limitations of Existing Studies*

Many of the prior works in the field of emotion detection, especially in digital image processing and deep learning algorithm, often encounter several limitations. Most use simple models such as the VGG16 or DCNN, which while serving basic image categorization lack the ability to discern the patterns for capturing emotion features required for categorization. These models often face challenges with overfitting, especially when dealing with limited or imbalanced datasets, resulting in suboptimal generalization to unseen data.

Additionally, previous studies frequently lack comprehensive evaluations across diverse emotion categories or robust datasets, which limit their applicability to real-world scenarios. Computational inefficiency is another significant drawback, as some models require extensive computational resources but fail to deliver proportionally high performance. In addition, few advanced data augmentation techniques are used and the absence of an extensive focus on specific domains leads to failures to reach better accuracy and reliability. Such limitations justify the need for more sophisticated and flexible strategies, as addressed in this research study. The novelty of our work lies in achieving the highest results with an increased number of classes, enabling more comprehensive emotion and sentiment analysis from diverse facial expressions in graphic art, surpassing the limitations of previous studies with fewer emotion categories.

TABLE I. SUMMARY OF EXISTING STUDIES

| Ref | year | Model | Dataset | Classes | Results |
|---|---|---|---|---|---|
| [14] | 2018 | CNN | Twitter images | 4 | 90% |
| [15] | 2023 | CNN | famous CK+, FER2013, and JAFFE | 3 | 95% |
| [16] | 2022 | BERT,CNN | MVSA-Multiple and T4SA | 2 | 93% |
| [17] | 2024 | LSTM | MBTI | 5 | 92% |
| [18] | 2020 | ConvNet-SVMBoVW model, SVM | IMDb | 5 | 91% |
| [19] | 2024 | Capsule with Deep CNN and Bi structured RNN | Twitter data | 4 | 95% |
| [20] | 2024 | LSTM-BiLSTM,BCNN | MVSA | 3 | 92% |
| [21] | 2023 | CNN | CK+, FER2013, and JAFFE | 4 | 94% |
| [22] | 2019 | CNN,LSTM | IMDB,Google news dataset | 4 | 92% |
| [23] | 2021 | CNN,LSTM,KNN | 2018 FIFA world cup tweets | 2 | 92% |
| [24] | 2022 | LSTM-CNN+GloVe | Tweets dataset | 3 | 93% |

## III. METHODS AND ARCHITECTURES

The two areas of artificial intelligence and deep learning have prompted notable improvement in understanding and discriminating demanding patterns in digital image. This study proposed the models that are embedded in recognizing colorless images depicting diverse artistic facial expressions. It encompasses high complication preprocessing, architecture, and large dataset to provide robust and accurate results. The architectures for the employed models are respectively shown in Fig. 1, which shows the basic workflow of any typical DL models with multi-tier structures of the architectures. Firstly, known as Dataset Collection, the image data pertinent to the process is accumulated. Then, Data Preprocessing is done to remove all irrelevant or duplicate data and adjust the data format for the model. The next step is Training with Models such as Inception v3 and ResNet50, which are deep convolutional neural network, which we extract features from images and learn some patterns from them. Fully connected layers follow the training process to combine the features which the model has learned with for the purpose of classification. It is then classified under different categories of different models, having considered the learned patterns. Moreover, the figure shows that Subtractive operations like Activation Function, Pooling, Flattening, Reduction of Overfitting and Optimizer are required to enhance the efficiency of the model and to avoid overfitting and thus more generalized results.

### *A. Dataset Preparation and Preprocessing*

Dataset consists of 32,298 grayscale images illustrating facial expressions and depicted as sketches with lead pencils. It includes seven emotion classes: This means the emotions, which are depicted in the images can be categorized into seven classifications, which are being angry, disgust, fear, happy, sad, surprise, and a neutral category. The challenges of the artistic and grayscale pictures as well as size and variability of the dataset are countered well, thus allowing the use of deep learning models.

The input data also handles through some manipulations to improve its compatibility with the selected deep learning models and more importantly to ensure that the models undergo stable training by applying two major steps of resizing and normalization. By changing the dimensions of the images to 299 by 299 pixels because that is the expected input size for model. There will be occasional variability in light conditions therefore the image’s data are normalized at a mean of [0.5, 0.5, 0.5] and standard deviation [0.5, 0.5, 0.5].

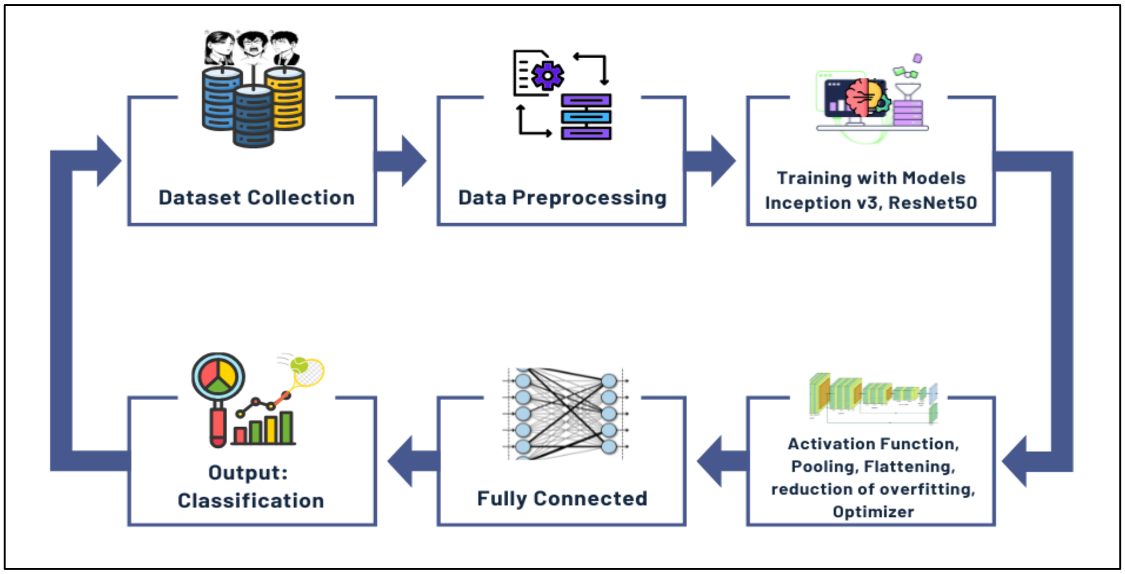

Fig. 1. Basic framework of any typical deep learning model.

These preprocessing steps assist in controlling fluctuations in the training curve using pixel values for normalization by following Eq. (1), there by having an optimized training process with good features. Table II contains comprehensive details of all symbols that are used in equations for better understanding.

$$P_{normalized} = \frac{P-\mu}{\vartheta} \qquad (1)$$

## B. Applied Deep Learning Models

In sentiment analysis method containing the feature extraction and the classification, both components are significant to the interpretation of the emotional context visually transferred. For example, in feature extraction, model captures the features such facial expressions, patterns, or textures as a representation of happy, sad, angry, and the like. Such features are embedded into the feature space of higher dimensions thus capturing relevant visual details. During this stage, these features extracted are then subjected to fully connected layers that try to map the image to sentiment categories. It makes it possible to capture slight changes in the expressions or other artistic features which are valuable to make robust sentiment categories even in highly diverse image sets.

*1) Inception v3 architecture:* Inception v3 is a deep learning architecture which optimizes computational speed and uses high effectiveness, as architecture shown in Fig. 2 for this study.

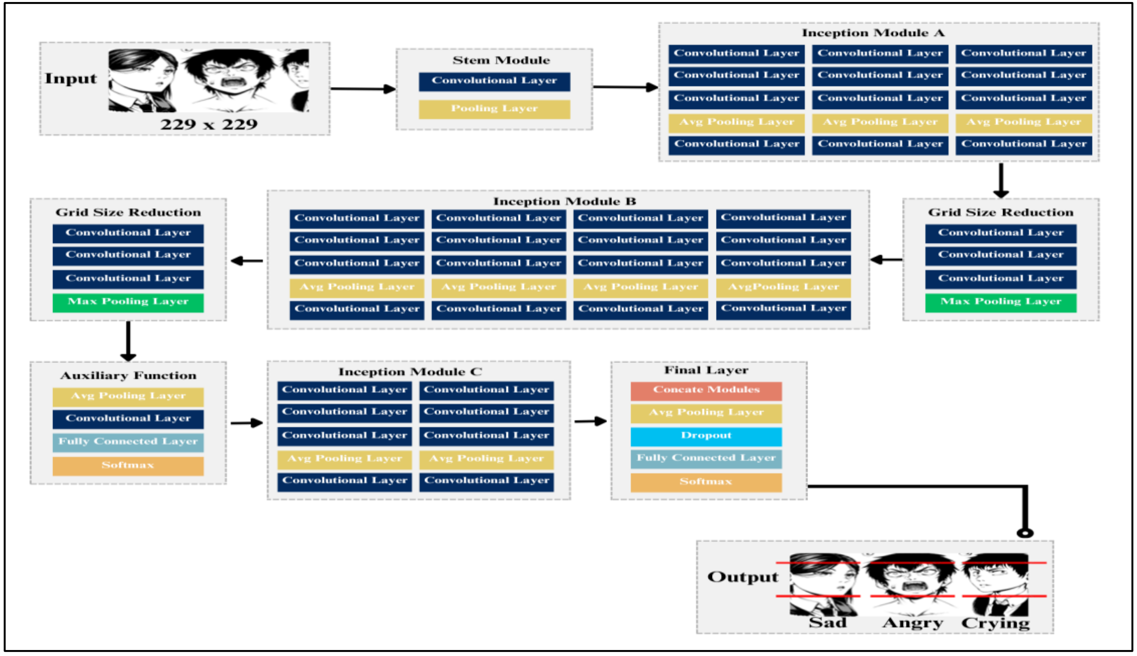

Fig. 2. Architecture of Inception v3 model.

This is done with the help of the inception module, which takes the input data and splits it through a series of channels every of which processes the data differently while trying to capture as many attributes of the input as $X \epsilon \mathbb{R}^{H*W*D}$ where $H$ is the height, $W$ is the width and $D$ is the depth showing number of channels. These paths are then added to produce a single output that can be represented as $O_{inception} \epsilon \mathbb{R}^{H\prime*W\prime*D\prime}$; thus, the network can capture a diverse set of features at multiscale level. To minimize computational complexity, certain design strategies have been applied, such as the factorized convolution, where a complex convolution is replaced by a series of simpler steps, as well as the dimensionality reduction using 1x1 convolutions, computed as in Eq. (2).

$$O_{inception} = concat(conv_{1*1}(X), conv_{3*3}(X), conv_{5*5}(X), conv_{1*1(3*3)}(X) \quad (2)$$

Where:

$$conv_{1*1}(X) = W_1 * +b_1, with\ W_1 \ \epsilon\ \mathbb{R}^{1*1*D*D_1} \text{ and } b_1 \ \epsilon\ \mathbb{R}^{D_1}$$

$$conv_{3*3}(X) = W_3 * X + b_3, with\ W_3 \ \epsilon\ \mathbb{R}^{3*3*D*D_3}$$

$$conv_{5*5}(X) = W_5 * X + b_5, with\ W_5 \ \epsilon\ \mathbb{R}^{1*1*D*D_5}$$

The auxiliary head is a technique of regularization in which gradients are added in the actual backpropagation processes during the training. The auxiliary classifier applies function over the featured space $\hat{y}_i^{auxiliary}$ produced by a certain layer of the network, computed as in Eq. (3).

$$L_{auxiliary} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i^{auxiliary}) \quad (3)$$

Also, Inception v3 requires Global Average Pooling (GAP) to decrease the size of feature maps $F \epsilon \mathbb{R}^{H*W*D}$ along with depth dimensions $D$ for training model and enhancing generalization to each feature map $f_d$, computed as in Eq. (4).

$$GAP(F) = \frac{1}{H*W} \sum_{i=j}^{H} \sum_{j=1}^{W} F_{ij}^{d} \quad (4)$$

Finally, the output is classified using activation function layer combines both main classification and the auxiliary loss, computed as in Eq. (5), which indeed makes the model very effective for large scale image recognition.

$$L_{total} = L_{main} + \delta L_{auxiliary} \quad (5)$$

*2) ResNet 50 architecture:* ResNet-50 is a categorized deep convolutional network model resolving vanishing gradients issue in very deep learning networks, as working defined in Fig. 3.

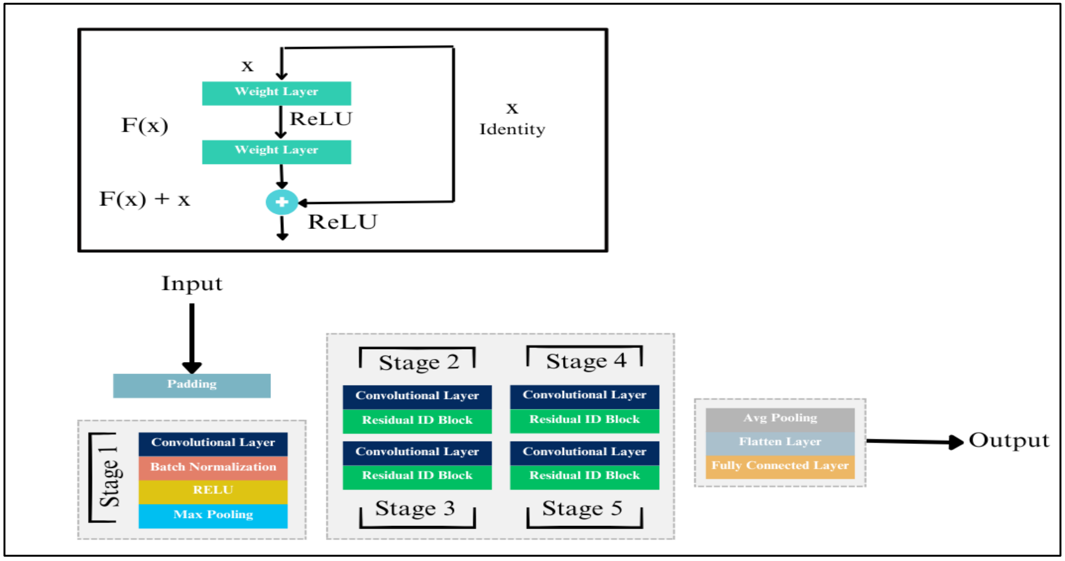

Fig. 3. Architecture of ResNet50 model.

With 50 layers and each layer of residual blocks. ResNet design is based on the concept of Residual block or skip connections that can bypass one or more layers of computations to provide the network with a method of training from scratch deeper networks that causes less degradation, at least theoretically. These skip connections are useful in reducing the vanishing gradient problem because they enable the gradients to flow directly through them using learning function $\mathcal{F}_k$ which is calculated as in Eq. (6).

$$y_k = \mathcal{F}_k\left(x_{k,}\{W_{k,i}\}\right) + x_k \quad (6)$$

The ResNet-50 model has been designed mostly for relatively small image classification problems and uses batch normalization and ReLU activation for enhanced results, computed as in Eq. (7).

$$z_k = \sigma\left(BN\left(\mathcal{F}_k\left(x_{k,}\{W_{k,i}\}\right) + x_k\right)\right) \quad (7)$$

As for the architecture, it is made up of the first convolutional layer, that is several stages of residual blocks where each block is made up of a 1x1 convolutional layer, a 3x3 convolutional layer and a final second 1x1 convolutional layer, defining the gradient flow using loss function $\mathcal{L}$, computed as in Eq. (8).

$$\frac{\partial\mathcal{L}}{\partial x_k} = \frac{\partial\mathcal{L}}{\partial y_k} + \frac{\partial\mathcal{L}}{\partial z_k} \cdot \frac{\partial z_k}{\partial x_k} \quad (8)$$

The architecture makes it easy to learn both low- and high-level features and that is the reason why this model is widely used in many computer vision tasks such as object detection and segmentation.

TABLE II. SYMBOLS DESCRIPTION OF APPLIED EQUATIONS

| Symbols | Description |
|---|---|
| $P - \mu$ | Mean values |
| $\vartheta$ | Standard deviation |
| $P$ | Original pixel values |
| $*$ | Convolution operation that are concate with final output |
| $y_i \epsilon \mathbb{R}^C$ | One hot-encoded true label vector for class $i$ |
| $\hat{y}_i^{auxiliary} \epsilon \mathbb{R}^C$ | Predicted probability distribution from the auxiliary classifier |
| $C$ | Number of output classes. |
| $F_{ij}^d$ | Feature value at position $(i, j)$ in the $d - th$ channel of the feature map |
| $\delta$ | Weight factor that balances the auxiliary loss relative to main loss |
| $y_k$ | Output of the $k - th$ residual block |
| $x_k$ | Input to the $k - th$ residual block |
| $\mathcal{F}_k\ (x_{k,}\{W_{k,i}\})$ | Learned residual function with weight $W_{k,i}$ |
| $z_k$ | Output of the Batch Normalization (BN) and activation |
| $\sigma$ | RELU activation function |
| $\frac{\partial \mathcal{L}}{\partial_{x_k}}$ | Gradient of loss with respect to input $x_k$ |
| $\frac{\partial \mathcal{L}}{\partial_{y_k}}$ | Gradient with respect to output $y_k$ |
| $\frac{\partial \mathcal{L}}{\partial_{z_k}}$ | Gradient of loss with respect to BN output $z_k$ |
| $\frac{\partial_{z_k}}{\partial_{x_k}}$ | BN output $z_k$ depends on the input $x_k$. |
| $TP\ and\ FP$ | True Positive and False Positive |
| $TN\ and\ FN$ | False Positive and False Positive |

## C. Performance Measures

To fully assess the performance of models, standard assessment metrics are utilized including accuracy, precision, recall and F1-score. They include information on correct classified instances, and are useful in cases where classes are imbalanced, or misclassifications to different classes cost differently. Accuracy is the simplest of all the performance measurement metrics that give the percentage of correct prediction of instances to the entire instances. But it might not be that effective in handling those datasets with imbalanced classes. Precision becomes important due to this, together with recall. Precision is the measure of the accuracy of the positive predictions calculated as the proportion of true positives to the total positive predictions and the false ones, it is valuable in those application domains where false positives carry serious implications (e.g., medical diagnoses). Recall or Sensitivity is equal to the relation of true positive findings to the sum of true positives and false negative results, which highlights the ability of the model not to miss any relevant cases. F1-score defined as the harmonic mean of precision and recall, is a valuable supplement to these two values, but most important when both measures are significant.

The training and validation accuracy and loss are important parameters to decide about the learning progress of a model. The accuracy of training measures the capability of the model on the training dataset, while the validation accuracy tests the model on how well it can perform on new dataset. The same about training loss that estimates the error during the learning process of the model on the training set, and validation loss that estimates the error on the validation dataset. Training loss should be decreasing while validation loss should be a plateau or on an increasing trend if there is not much data for training or training data is limited rather than showing a decreasing trend having a low value is best for a well-generalized model. Thus, Table III indicates the metrics employed to adjust the models depending on the context of evaluation, which will be highly beneficial for practitioners.

TABLE III. EVALUATION PERFORMANCE MEASURES

| Sr. No | Metrics | Equation | Purpose |
|---|---|---|---|
| 1 | Accuracy | $\frac{TP+TN}{TF+FN+FP+TP}$ | Measures overall correctness |
| 2 | Precision | $\frac{TP}{TP+FP}$ | Focus on avoiding false alarms |
| 3 | Recall | $\frac{TP}{TP+FN}$ | Emphasizes capturing all actual positives |
| 4 | F1-score | $\frac{2(Precision*Recall)}{Precision+Recall}$ | Indicates overall performance balance |

# IV. RESULTS AND DISCUSSION

The findings on how facial emotions are classified using the deep learning pre-trained models, Inception v3 and ResNet50 model can be a useful guide on the efficiency of deep learning for facial emotions classification, as shown in table IV. The main aim of the study is to distinguish between emotions like 'angry," 'crying', 'embarrassed' , 'happy', 'pleased,' 'sad,' and 'shock' through facial expression, and the results of the study can be given multiple interpretations.

TABLE IV. RESULTS OF MODEL PERFORMANCE ACROSS ALL MEASURES

| Models | Training | | | | Validation | | | |
|---|---|---|---|---|---|---|---|---|
| | *Accuracy* | *Precision* | *Recall* | *F1-Score* | *Accuracy* | *Precision* | *Recall* | *F1-Score* |
| ResNet50 | 98 | 98 | 98 | 98 | 67 | 66 | 67 | 65 |
| **Inception v3** | **99** | **99** | **99** | **99** | **76** | **78** | **76** | **76** |

## A. Hyperparameter Settings

The configurations of the model are adjusted to improve its performance for the emotion classification task, as shown in Table V. An optimizer learning rate is used to adjust generalization of the pre-trained model while practicing on the dataset without causing much alteration of the learnt parameterization. The number of batches has been defined in the manner to optimize the computational resources and to maintain steady gradients during the training phase. This algorithm is used due to its adaptive learning rate for each parameter what makes possible to obtain faster convergence with better generalization. Like the previous optimizers, this optimizer does not require specific parameter settings since it utilizes standard settings for its operation for efficient weight updates during the optimization process.

TABLE V. VALUES OF HYPERPARAMETERS

| Parameters | Values |
|---|---|
| Learning Rate | 0.0001 |
| Batch Size | 32 |
| Optimizer | Adam |
| Adam Settings | Lr = 0.0001, beta1 = 0.9, beta2 = 0.999 |
| Loss Function | Cross Entropy Loss |
| Epochs | 30 |
| Model Architecture | Inception v3, ResNet50 |

For the loss function of this task, Cross Entropy Loss was adopted due to its application to a typical multi-class classification as it combines the softmax activation function and negative log-likelihood loss optimally. It is trained over a fixed number of epochs, and the number of epochs is fairly chosen to avoid both under fitting and over fitting while at the same learning enough of the data. Thus, the model architecture contains inception modules which in a way sample across scales using multiple different spatial convolutional features. Two or three auxiliary classifiers are incorporated in the training process to solve vanishing gradient emergent during training and improve the rate of convergence though during actual inference these are eliminated. Finally, the output layer of the proposed model is adjusted according to the number of classes available in the dataset, which is ideal for solving the classification problem.

### B. *Results with ResNet50 Model*

First, the training results include high accuracy of 98.18% and low training loss of 0.0401 and a very high precision, recall, and F1-score all of which are 98.18%, so the training signals have been well learnt by the model. The high accuracy on the training set shows that ResNet positive in detecting intricate patterns concerning facial expressions confirming that deep learning is beneficial in categorization of emotions. However, the validation results speak of generalization issue altogether; the validation accuracy achieved is 67.03 % with a validation loss of 1.1136, and quite low precisions with 66.98 %, recall with 67.03%, F1-score with 65.76%. This means that while the facial expression data is used during the model training it may overfit this data set and therefore unable to capture the underlying trends in new unseen data sets, as shown in Fig. 4. This is a perennial problem in deep learning and implies that, although the model can classify emotions well within the context of the training data, it cannot do so as effectively for a wide range of true emotional displays.

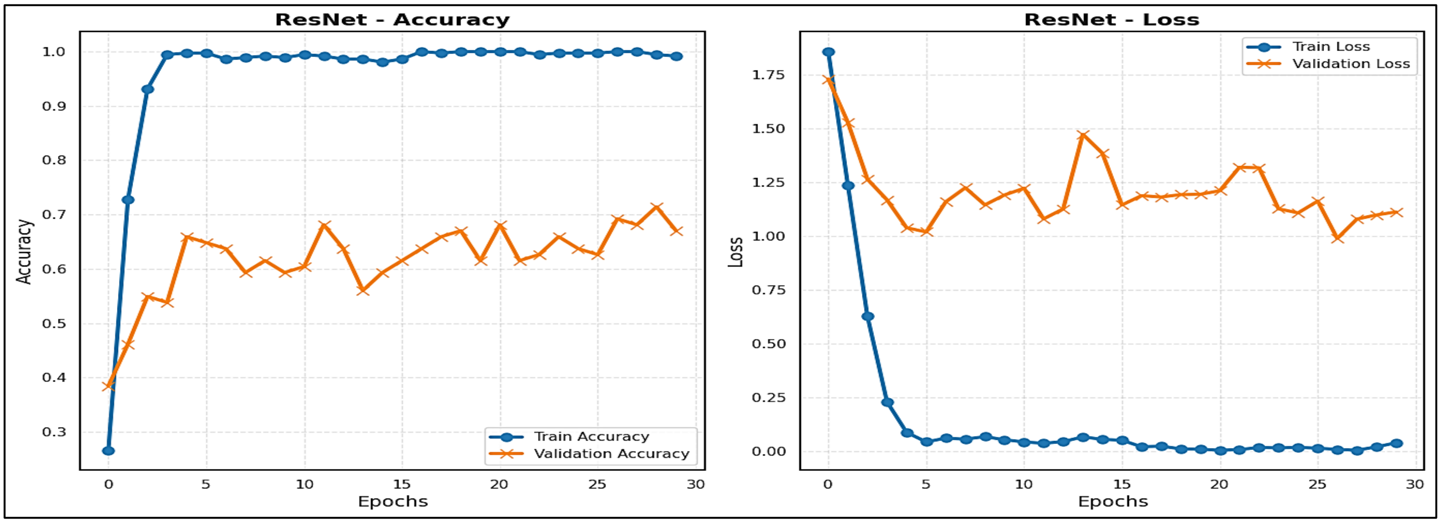

Fig. 4. Analysis of ResNet50 model performance across accuracy and loss.

Following that is the confusion matrix as shown in Fig. 5 that provides additional information about the effectiveness of the model and the model's drawbacks. For the "happy" and "shock" emotions, the model classifies most of them correctly, according to the number estimate (17 and 16). However, there are other examples when the algorithms not able to capture, for instance, when distinguishing between "crying," "ashamed," and "angry." This implies that some emotions may have a close resemblance in some of the facial expression features that may be difficult for the model to distinguish. Especially, the examples like "crying" or "embarrassed" are less distinguishable for the model since it must differentiate between more shades of the mentioned feelings.

### C. *Results with Inception v3 Model*

Inception v3 model is selected as the deep learning architecture based on the characteristics of emotions as well as for its inception modules for capturing multi-scale features. The graph presenting in Fig. 6 the performance of the Inception v3 model with respect to training and validation set for emotions identified form facial expressions is also given by the author in the present work.

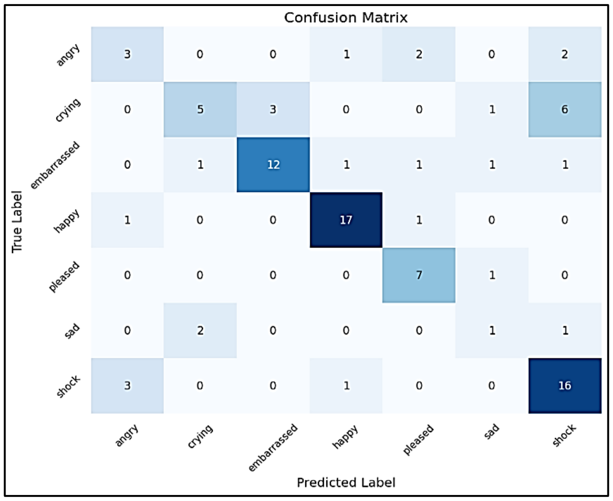

Fig. 5. Confusion matrix showing sentiment using ResNet50.

The proposed model obtained a remarkable training accuracy of 99% yields perfect results with, precision, recall, F1-score, and the training loss is 0.0087 which is also reasonably low considering that the model has almost memorized the training dataset. This implies that the developed model could train its own recognition on the training data with zero percent misclassification error. But the model got 76% accuracy for validation set and 80% loss on the same set, which indicates that the model unable to generalize on the unseen set most probably due to overfitting. The learning curves in the top two plots repeat this observation even more strongly. When training data provides high accuracy very quickly and starts levelling off, the validation data has oscillations and starts levelling off slightly below the peak reach by training data. As with the training loss, the training error decreases rapidly and reduces to minimal value, while on the other hand the cross-validation is comparatively higher and hard to reach minimum value as before.

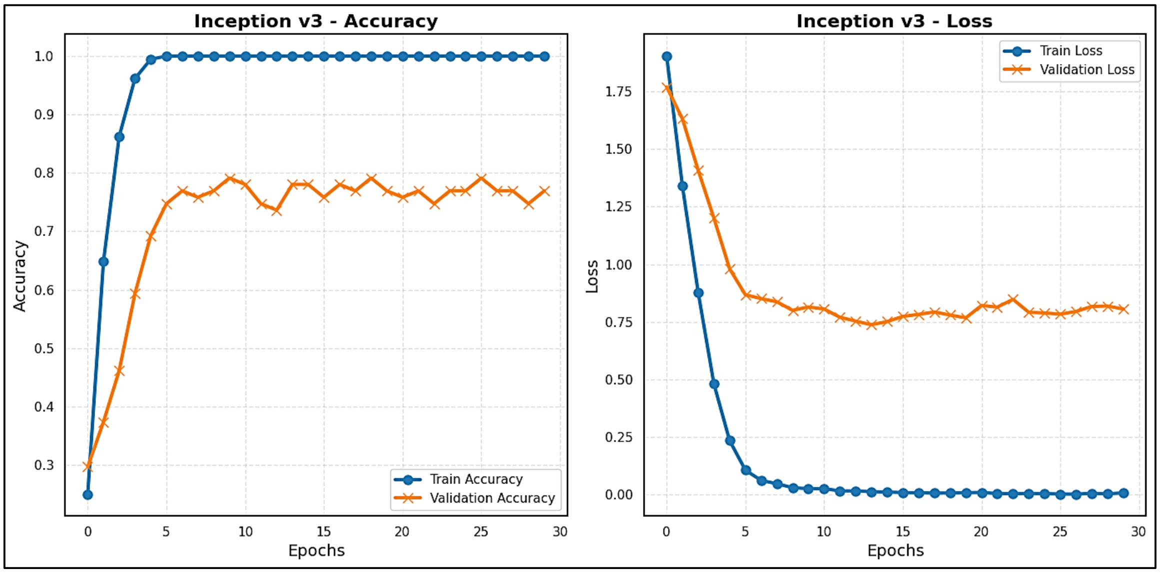

Fig. 6. Analysis of Inception v3 model performance across accuracy and loss.

Confusion matrix as shown in Fig. 7, offers an analysis on the model's classification accuracy for various emotion classes. Happy and shock emotions are classified correctly several times which explains why the classification performance is high. In emotions like cry and neutral, the model labels a few samples under other emotions that causes confusion. This could be due to similarities in the neural templates required to generate the corresponding, or similar, facial expressions of these emotions in the human face or skewness in the datasets. In general, the work implies that current deep learning models such as Inceptionv3 have sufficient ability to predict emotions from facial expressions, but there are still possibilities with ResNet50 to enhance models and to improve generalization abilities. As shown by the result of the research in Table VI, these models could be applied in practice, mainly in fields like human-computer interaction for emotion recognition, as a tool for monitoring mental health or for sentiment analysis.

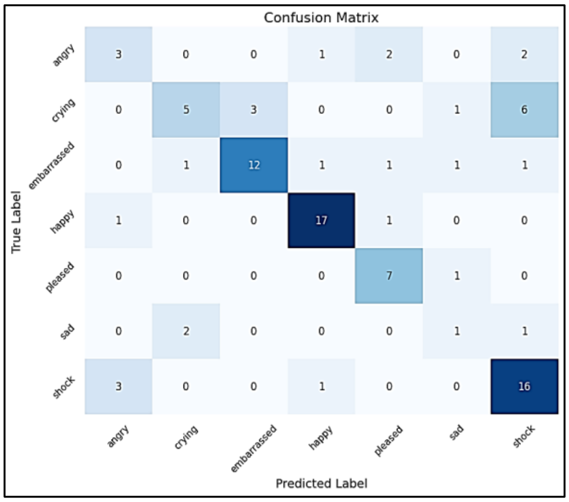

Fig. 7. Confusion matrix showing sentiment using Inception v3.

TABLE VI. RESULTS OF MODEL ACCURACY AND LOSS SUMMARY

| **Models** | **Training** | | **Validation** | |
|---|---|---|---|---|
| | ***Accuracy*** | ***Loss*** | ***Accuracy*** | ***Loss*** |
| ResNet50 | 0.98 | 0.04 | 0.67 | 1.13 |
| **Inception v3** | **0.99** | **0.008** | **0.76** | **0.80** |

### D. *Comparison with Existing Studies*

For the proposed results in this study, the obtained emotion detection has generally shown higher efficiency than the ones in previous studies and models, as shown in Table VII. Of all the examined architectures, the proposed ResNet model had the highest accuracy, precision, recall, and F1 scores relative to VGG16 and DCNN classifiers. Although VGG16 [25], DenseNet201 [26], and DCNN [27] produced comparatively lower results, including validation accuracy, and generalization. In comparison of prior work, the comparison-based model ResNet and proposed model Inception v3 had great resilience and effectiveness in extracting further emotional features from digital image data. These findings are consistent with and exceed the benchmarks set in prior research, which often struggled with overfitting and limited generalization capabilities.

TABLE VII. COMPARISONS OF RESULTS WITH EXISTING STUDIES

| **Ref** | **Year** | **Model** | **Dataset** | **Results Acc (%)** |
|---|---|---|---|---|
| [25] | 2020 | VGG16 | Media Art | 42 |
| [26] | 2021 | DenseNet201 | Custom Dataset | 35 |
| [27] | 2022 | DCNN | Artificial Images Data | 39 |
| **Proposed** | **2024** | **ResNet** | **AI vs Human** | **98** |
| | | **Inception v3** | | **99** |

Fig. 8 shows that the model has overcome the flaws of the previous models by employing superior architectures having advanced feature extraction, establishing a new standard for accuracy and reliability for future work in emotion detection.

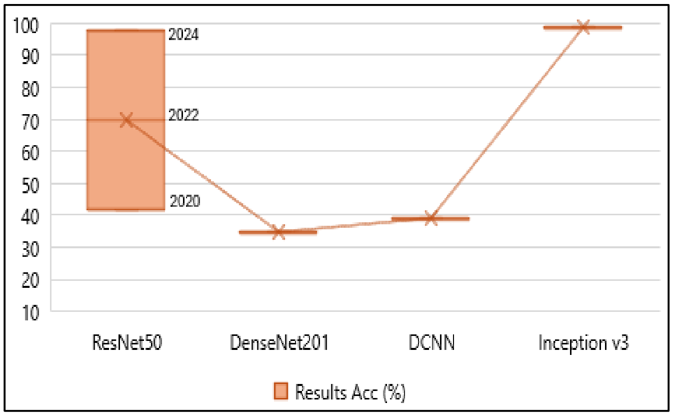

Fig. 8. Comparative analysis of proposed models with existing studies.

## V. CONCLUSION

The extensive development in AI has significantly transformed different fields such as emotion detection, digital image processing, and facial expressions analysis. Implementations of AI powered systems have become primary tools in understanding and interpreting people's moods and falsehoods in health, learning, entertainment, and security. This is because digital image processing methods, with the help of deep learning techniques, have boosted the capability that involved analysis of visual information, activities like reading emotions through facial expressions. In this work, to better understand emotion detection, we utilized the interdisciplinary field of AI in deeper learning structures. Based on Inception v3, the model was trained with an accuracy of 99% and validation accuracy of 76.92%. Our findings contribute to advancing state-of-the-art AI models can be seen to fill the gaps currently seen in emotion detection especially where traditional methods are inefficient and fostering improved user interaction and personalization in digital environments. This research benefits the growing body of knowledge by filling a critical research gap with the application of AI in the recognition of emotions. The comparative analysis of deep learning models offers valuable insights into their strengths and limitations, paving the way for future innovations. Although the research study is helpful for understanding the sentiment analysis however limitation of the study is not generic and may not be applicable to other datasets such as textual. Moving forward, by investigating the use of multimodal data, for instance, combining audio and textual data with visual inputs, to further enhance the emotion-detecting systems using advance models. Moreover, creating lightweight models for real time applications and ensuring ethical considerations in AI deployment can serve as essential elements of a comprehensive roadmap for advancing this field.

## REFERENCES

[1] Sudha, K., Muthumarilakshmi, S., Kavitha, G., Hashini, S. and Kumar, V.N., 2023, October. Sentiment Analysis on Text data: Methods, Applications, Challenges and Future Directions. In 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT) (pp. 1-7). IEEE.

[2] Mahmood, A., Khan, H.U. and Ramzan, M., 2020. On modelling for bias-aware sentiment analysis and its impact in Twitter. Journal of Web Engineering, 19(1), pp.1-27.

[3] Iqbal, S., Khan, F., Khan, H.U., Iqbal, T. and Shah, J.H., 2022. Sentiment analysis of social media content in pashto language using deep learning algorithms. Journal of Internet Technology, 23(7), pp.1669-1677.

[4] Mutanov, G., Karyukin, V. and Mamykova, Z., 2021. Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. Computers, Materials & Continua, 69(1).

[5] Ahmad, W., Khan, H.U., Iqbal, T. and Iqbal, S., 2023. Attention-based multi-channel gated recurrent neural networks: a novel feature-centric approach for aspect-based sentiment classification. IEEE Access, 11, pp.54408-54427.

[6] Li, X. and Li, Y., 2024. Deep Learning and Natural Language Processing Technology Based Display and Analysis of Modern Artwork. Journal of Electrical Systems, 20(3s), pp.1636-1646.

[7] Rane, N., Choudhary, S. and Rane, J., 2024. Artificial intelligence, machine learning, and deep learning for sentiment analysis in business to enhance customer experience, loyalty, and satisfaction. Available at SSRN 4846145.

[8] Mao, Y., Liu, Q. and Zhang, Y., 2024. Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University-Computer and Information Sciences, p.102048.

[9] Yuan, J., Mcdonough, S., You, Q. and Luo, J., 2013, August. Sentribute: image sentiment analysis from a mid-level perspective. In Proceedings

of the second international workshop on issues of sentiment discovery and opinion mining (pp. 1-8).

[10] Liu, H., Chatterjee, I., Zhou, M., Lu, X.S. and Abusorrah, A., 2020. Aspect-based sentiment analysis: A survey of deep learning methods. IEEE Transactions on Computational Social Systems, 7(6), pp.1358-1375.

[11] Ahuja, G., Alaei, A. and Pal, U., 2024. A new multimodal sentiment analysis for images containing textual information. Multimedia Tools and Applications, pp.1-30.

[12] Baldoni, M., Baroglio, C., Patti, V. and Schifanella, C., 2013. Sentiment analysis in the planet art: A case study in the social semantic web. New Challenges in Distributed Information Filtering and Retrieval: DART 2011: Revised and Invited Papers, pp.131-149.

[13] Pathak, A.R., Pandey, M. and Rautaray, S., 2021. Topic-level sentiment analysis of social media data using deep learning. Applied Soft Computing, 108, p.107440.

[14] Kumar, A. and Jaiswal, A., 2018. Image sentiment analysis using convolutional neural network. In Intelligent Systems Design and Applications: 17th International Conference on Intelligent Systems Design and Applications (ISDA 2017) held in Delhi, India, December 14-16, 2017 (pp. 464-473). Springer International Publishing.

[15] Meena, G., Mohbey, K.K., Kumar, S., Chawda, R.K. and Gaikwad, S.V., 2023. Image-based sentiment analysis using InceptionV3 transfer learning approach. SN Computer Science, 4(3), p.242.

[16] Ghorbanali, A., Sohrabi, M.K. and Yaghmaee, F., 2022. Ensemble transfer learning-based multimodal sentiment analysis using weighted convolutional neural networks. Information Processing & Management, 59(3), p.102929.

[17] Naz, A., Khan, H.U., Alesawi, S., Abouola, O.I., Daud, A. and Ramzan, M., 2024. AI Knows You: Deep Learning Model for Prediction of Extroversion Personality Trait. IEEE Access.

[18] Kumar, A., Srinivasan, K., Cheng, W.H. and Zomaya, A.Y., 2020. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. Information Processing & Management, 57(1), p.102141.

[19] Islam, M.S., Kabir, M.N., Ghani, N.A., Zamli, K.Z., Zulkifli, N.S.A., Rahman, M.M. and Moni, M.A., 2024. Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. Artificial Intelligence Review, 57(3), p.62.

[20] Fang, Y. and Wang, Y., 2024. Cross-modal Sentiment Analysis of Text Image Fusion Based on Hybrid Fusion Strategy. Informatica, 48(21).

[21] Meena, G., Mohbey, K.K. and Kumar, S., 2023. Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach. International journal of information management data insights, 3(1), p.100174.

[22] Rehman, A.U., Malik, A.K., Raza, B. and Ali, W., 2019. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. Multimedia Tools and Applications, 78, pp.26597-26613.

[23] Hegde, S.U., Zaiba, A.S. and Nagaraju, Y., 2021, February. Hybrid cnn-lstm model with glove word vector for sentiment analysis on football specific tweets. In 2021 international conference on advances in electrical, computing, communication and sustainable technologies (ICAECT) (pp. 1-8). IEEE.

[24] Ismail, A.A. and Yusoff, M., 2022. An efficient hybrid LSTM-CNN and CNN-LSTM with GloVe for text multi-class sentiment classification in gender violence. International Journal of Advanced Computer Science and Applications, 13(9).

[25] Salmaneunus (2020) Image classification in pytorch: Manga Facial Expression, Kaggle. Available at: https://www.kaggle.com/code/salmaneunus/image-classification-in-pytorch-cifar10 (Accessed: 02 December 2024).

[26] Stpeteishii (2021) Manga Face DENSENET201, Kaggle. Available at: https://www.kaggle.com/code/stpeteishii/manga-face-densenet201 (Accessed: 02 December 2024).

[27] Vishalkalathil (2023) Manga facial expression classifier DCNN, Kaggle. Available at: https://www.kaggle.com/code/vishalkalathil/manga-facial-expression-classifier-dcnn (Accessed: 02 December 2024)