*Journal of Computer Science*

Research Article

# A Sentiment Analysis Approach for Affective State Recognition With SWISH-Activated DenseNet-169 in Human Event Forecasting

**Karthik Elangovan[1], Prabhakaran Paulraj[2], Dinesh Babu G L[3], A. S. Anakath[4] and Jayaraj Ramasamy[5]**

[1]*Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Vel Tech University, Chennai, India*
[2]*Department of Electrical, St Joseph University in Tanzania, Dar es Salaam, Tanzania*
[3]*Department of Artificial Intelligence and Machine Learning, Saveetha Engineering College, India*
[4]*Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, SIMATS, India*
[5]*Department of Computing, De Montfort University Kazakhstan, Kazakhstan*

*Article history*
Received: 25-06-2025
Revised: 12-03-2025
Accepted: 07-07-2026

**Corresponding Author:**
Prabhakaran Paulraj
Department of Electrical, St Joseph University in Tanzania, Dar es Salaam, Tanzania
Email: prabakaran.p@sjuit.ac.tz

**Abstract:** Detecting human feelings from images is perhaps one of the most useful and challenging social communication problems in research. Emotion detection-based deep computing outperforms classic image processing techniques. Facial recognition technology has permeated every aspect of our daily lives, from unlocking our phones to getting private file access on our systems. Machine Learning and Deep Learning make it easy to detect facial expression by using the method of sentiment analysis, which in turn makes it possible to identify the status of the emotions, which can be used to predict human emotions. ML has opened a new world of utilizing technology, from self-driving automobiles to recognizing faces on your mobile lock screen. Face recognition finds individuals in the landscape using AI, ML and DL algorithms. Further confirmations using big datasets with both positive and negative photographs, after all the facial traits have been recorded, aid in confirming that the image is truly of a human face. Face detection, feature extraction, and emotion categorization are all part of the basic facial emotion identification process. The user's facial expression is dynamically captured from a video stream or statically captured image, and the image is further processed. The seven human facial emotions are classified in the processed image by applying machine learning-based proposed SWISH DenseNet-169 image classifier thereby it facilitates the forecasting of human moves. In this work, SWISH DenseNet-169 is contrasted with cutting-edge techniques such as EfficientNet, CNN, Vision Transformers, RCNN, and LSTM. The proposed DenseNet-169 is enhanced the accuracy measures and it was executed on a variety of hyman expression data sets includes CK+, JAFFF, BES, EMOTI-W, and IAP.

**Keywords:** AI, DenseNet, Facial Emotion, Gini Index, ML, Sentiment Analysis, SWISH DenseNet, Transformers

## Introduction

A subfield of machine learning technology called sentiment analysis (Wang et al., 2024) replicates cognitive functions performed by humans using machines. Sentiment analysis uses a set of user generated data and images to analyze information patterns to perform cognitive decisions to correlate the information. A rational outcome arrives through sentiment analysis, which augments various forms of automation by harnessing neural networks (Khan et al., 2019), machine learning, and deep learning. Sentiment Analysis is an analytical technique which isolates the polarity of the input which can be text, speech, facial expressions, emotional expressions etc. (Cha et al., 2020). Sentiment analysis is concerned with the definition of emotional inclination of text by classifying it into such classes as positive, negative, or neutral. Likewise, opinion mining is a collection of

methods of analytic tools that are used to detect and comprehend subjective views and expressive emotions constituted in text. Sentiment analysis gives a general solution to the given input whereas Opinion mining (Tsao et al., 2010) gives an in-depth subjective response from the general solution obtained from sentiments. Humans can express their perspective through verbal and non-verbal mediums. Verbal medium includes views derived from speech and textual modes. Non-verbal Medium includes views obtained from facial expression, gestures, and Paralinguistics. The Effective Mode of recognizing emotions is through Facial expressions. Classification of Emotions through textual form gives a linear outcome which may result in lack of accuracy in the outcome whereas classifying emotions through facial expressions gives a non-linear outcome such as sarcastic expressions. We require a system to deal with non-linear expressions of emotions.

The proposed system converts the non-linear outcomes into linear outcomes. The SWISH DenseNet-169 algorithm is used for facial expression detection to identify human confrontations in the backdrop. After all the facial features (Wang et al., 2020) have been recorded, subsequent validation utilizing sizable datasets with both positive and negative photos helps to validate that the image is indeed of a human face. The main objective of this research is to predict the seven basic human facial emotions from non-verbal forms of data for finding mood of the person. The progress of Machine learning techniques has been challenging when it comes to Computer Vision, Image processing and Video processing. The Proposed SWISH DenseNet-169 can be used in various real time applications for detecting and classifying emotions. Facial Emotion recognition has become an essential part of the field of market research and healthcare where emotions can be used to define the original opinion of a product and to detect the seriousness of the illness for necessary medications respectively. Face capture, feature extraction and emotion categorization comprise the fundamental process of Facial Emotion Recognition (FER). We have developed a method that performs opinion mining (Rodrigues et al., 2020) using image data. In this instance, the DenseNet strategy was used, a novel neural network strategy. The aspiration of the presented approach is to accurately predict sentiment using facial data and yield a more accurate result than past techniques. This method increases the industries where the analysis can be applied.

In order to critically analyze the efficiency of the presented SWISH-activated DenseNet-169 structure in facial affect recognition, a number of already established baseline models were utilized and tested. This baseline methods are a collection of various paradigms of learning that include traditional convolutional learning, efficient deep architecture, sequence learning, and attention-based models. They can be included to provide a thorough comparative analysis and demonstrate the relative superiority of the suggested approach in the processing of complex expressions of facial emotions that are non-linear. The CNN-based models acquire local receptive field learning and parameter sharing, which are useful in low-level patterns, including edges, contours, and textures (Jaiswal et al., 2020). Nevertheless, conventional CNN models tend to be limited in feature reuse and degradation challenges as network depth grows and as such, they may be limited in their capacity to capture subtle and compound emotional signals within facial expressions. Another baseline model that was used to overcome computational efficiency and scalability was EfficientNet. EfficientNet employs a scaling strategy on compounds which is uniform that balances network depth, width and resolution so the network can perform better with fewer parameters (Yang et al., 2018). EfficientNet is highly efficient and accurate at image classification; it predominantly uses global scaling methods of EfficientNet to scale attributes for each class. This global approach lags the emotional variations within a localized context, whereas micro expression helps to define a person's affective state.

RCNN+LSTM networks are able to model temporal and contextual information, utilized cell-based architectures (Baddar et al., 2022) for extracting the features in conjunction with recurrent connections to connect the spatial regions within each other through the context of the overall output. LSTMs can be successfully applied in the context of sequential dependencies modeling and have been applied in assessing emotion-based patterns with facial data being processed as ordered feature representations. Nonetheless, both are more sequential/temporal oriented and might not be able to make the most of dense spatial feature propagation, needed by static facial emotion recognition. Vision Transformers were considered one of the transformer-based baselines because of their self-attention mechanisms, which are capable of long-range dependencies modeling. Vision Transformers are images divided into patches and handled on a global scale, which could capture holistic facial structures (Saxena et al., 2020). However, ViTs usually need large data sets to train effectively and can perform poorly in emotion recognition when not a lot of data is available because local facial expression changes like eyebrows movement or lips shape are important when analyzing affective expressions. The proposed SWISH-activated DenseNet-169 architecture proposes dense feature connectivity, which allows maximum reuse of features and propagate gradients effectively in any of the layers. SWISH activation functionality contributes to learn non-linear representations whereby, the model can learn to convert complex facial emotion configurations to discriminative linear representations. Consequently, the suggested SWISH DenseNet-169 exhibits high robustness, accuracy, and generalization ability as opposed to the baseline models, especially in the ability to detect fine and non-linear expressions of emotions based on facial images.

*Literature Review*

Wang et al. (2024) created an ECDNet to optimize the facial expression recognition. They isolated core expressive features of a face by separating them in a systematic manner out of other confounding facial qualities, including identity or lighting, found in an original image. The existing methods analyzed only facial expressions and depend on small amounts of facial attribute annotations. ECDNet reconstructs facial images by synthesizing and merging the facial appearance, facial geometric data features and disentangling muscle movements associated with expression from complementary facial features. The model employs methods of expression incentive and inhibition to target expression features to induce and suppress, respectively, to ensure the model pulls out precise, complete representations of both expression components and non-expression components. A comprehensive and rigorous evaluation of ECDNet against previous facial expression recognition methods was performed on the well-studied datasets RAF-DB, AffectNet, and CAER-S, demonstrating a significant accuracy advantage over a few state-of-the-art facial expression recognition methods. The model demonstrates improved accuracy over previous methods but also improved visualization of the nuanced facial muscle movements associated with emotion. ECDNet shows minimal deployment times as it retains and leverages information about what comprised the components of the model, indicating that it is efficient and capable of being scaled to other domains, such as human-computer interactions or remote education. While this study has developed deep learning models for expression disentanglement, others have explored alternative modalities such as bio signals to mitigate challenges of occlusion in virtual environments.

In an article by Cha et al. (2020), the researchers designed a facial expression recognition (FER) system using facial Electromyogram (fEMG) signals collected from facial electrodes placed around the eye that will physically be farthest from obstruction via a head-mounted display in social virtual reality environments. This method has theoretical benefits over conventional techniques, for example, dedicated cameras or strain sensors; dedicated cameras can have occlusion issues, as well, the costs related to attaching and calibrating cameras for facial expression tracking in virtual reality environments. Quality subpar cameras have difficulty detecting small articulations of micro expressions, and strain sensors are challenged by the difficulty of calibrating them to track the dynamic facial expression movement in real-time. The authors employ a Riemannian manifold-based classifier which greatly reduces the need for many labeled training samples for FER, without reducing classification performance. Furthermore, the authors run the evaluation study with 42 participants, achieving a mean classification of 85% accuracy for classifying 11 total unique facial expressions when variance had only single labeled training sample per expression class.

Fuad et al. (2021) summarized various aspects about the dramatic transformation in face recognition techniques due to the emergence and widespread use of deep learning technology. Specifically, they discussed the evolution of popular current face recognition techniques and offer comparisons amongst these techniques, including Convolutional Neural Networks (CNNs), deep reinforcement learning, autoencoders, and generative adversarial networks. The authors also gave a survey of the 171 systems examined for face recognition, specifically highlighting the algorithm structural differences in each system as well as the differences in loss functions and activation functions, and the base data sets. This paper also discussed the issues of pose, age, illumination, occlusion, and make-up related differences that face recognition systems have to deal with. Furthermore, it describes the advancements from classical statistical and machine learning methods to deep learning methods, which are achieved as near to 100% accuracy on public datasets as possible. Fuad et al. concluded with a vision of productive future trends and open problems for deep learning face recognition research, emphasizing the collection of larger and more variegated datasets for training to provide performance improvements and go towards robustness and generalization in the wild.

The earlier works have some shortfalls to identify human faces in a photograph that emphasizes environmental and structural variation for face recognition. This difficulty in a photograph is typically derived from the human face versions, that involves occlusions, head poses (off-aircraft rotations), face tilting (in-aircraft rotations), illumination, facial expressions, and scales. Also, environmental problems which include lighting fixtures conditions, photograph quality, and cluttered backgrounds. This sets out an approach Tsao et al. (2010), that is based on the MAFIA algorithm, which could be the next generation of face detection technology. The two stages of the suggested method are detection and education. During the education phase, they transformed each education photograph into a part photograph by employing Sobel's part detection operator, morphological operator, and thresholding. A sliding window is used in the detection phase for analyzing images on distinct scales. Every sliding window is considered to have an individual's visage if it passes the face detector. The suggested method can consistently identify the distinctive styles that highlight the most prominent face characteristics. While the SNoW technique performs first-class inside the BioID dataset, the suggested technique performs first-class on the MIT-CMU dataset.

Building on these foundations, researchers began to diverge and focused on deep learning-based emotion detection frameworks that led to significant improvements upon existing traditional image-processing techniques.

Deep learning primarily based totally emotion detection offers overall performance higher than conventional strategies with picture processing (Jaiswal et al., 2020). The layout of a synthetic intelligence device able to detect emotion via facial expressions. They talk about the manner of emotion detection, which incorporates essentially 3 fundamental steps: Face detection, capabilities extraction, and emotion classification. In a somewhat similar capacity, the demand for sustainable authentication purple in information-rich environments raised consideration of using CNN-based recognition systems in smart classrooms (Khan et al., 2019). One of the predominant problems in coping with massive quantities of records is authentication and face reputation is taken into consideration as a dependable solution. They have proposed a set of rules that establishes a clever lecture room for students' attendance and the usage of face reputation. The proposed machine performed 97% accuracy on validation and trying out records.

In the same spirit of class-room applications, others explored lightweight and mobile solutions (Ghali et al., 2018) for emotion recognition. This described emotion recognition for digital images using an android app and detected seven basic facial emotions. Open-CV and Fisher face algorithm combines the PCA and LDA algorithms. Interactive system that detects facial expression is performed through comparisons of features of eye and mouth coordination of an image database to identify similar patterns of expression. The implementation of this emotional intelligence in the computing system renders them more responsive to the user. The outcome increased in the quality of human-machines cooperation, which results in more natural, satisfying and fruitful interactions.

Building on facial recognition work, set the stage in the larger context of big data analytics thinking about emotion and opinion mining. The technological advancements characterized by the connectivity and widespread use of mobile devices have expanded the volume of data which increases the penetration of the internet in all facets of contemporary society. The volume of unstructured data created, comprised of photographs, videos, and mostly text strings of posts on social networks, and is a main driver of this volume. Social media data is complicated, and we cannot rely on existing tools to find genuinely actionable insights. We must create entirely new approaches, this persistent and perpetual dilemma that drives a lot of innovation in data sciences and big data analytics. The study of Rodrigues et al. (2020) investigates customer experience discovery through the opinion mining framework using unstructured data that was able to apply to or relevant the realities of small and medium sized business.

In addition to data utilization strategies, other researchers explored improving recognition for emotions through advances of feature extraction and classification techniques called Vector Field Convolution (VFC) (Hazar et al., 2013). The facial expressions depth varies from one person to any other which makes it tough to outline green norms and trendy regulations to efficiently classify human emotions. It's far targeted in this tough hassle to outline a distinguished and applicable set of prediction regulations capable of understanding facial expressions regardless of emotion depth variation. The VFC version supplied an achievement of 81.42% with five instructions and 75.32% with 7 instructions. Hazar et al. (2013) cognizance on distinct strategies of expression recognition. First template-primarily based totally technique and then appearance-primarily based totally technique called precept thing analysis. In the proposed VFC study have explained a unique method for expression detection of an individual. Expression is one of the crucial elements of humans. In VFC records are used by the records mining statistical method, the usage of statistics advantage and gini index on the way to understand the appropriate emotion and the proposed VFC attained 80.9 % accuracy.

However, as Choi et al. (2018) pointed out, many holistic feature-based methods have shortcomings with respect to distortion in the real world, spurring exploration in improvements for discriminating analysis techniques. The holistic function-primarily based totally strategies carry out favorably for normalized face pics received in restrained surroundings along with the inner laboratories. The variety of accessories added to the different versions of the human face is substantially decreased via a huge distortion of the entire face image. A new composite feature vector construction technique for face recognition was developed by these authors through the use of discriminant analysis. Their technique outperformed existing baseline models (i.e., Eigenfaces and Fisherfaces) by 3%.

After these improvements, greater attention turned to social media sentiment analysis, with hybrid neural networks suggesting robust paths for resolving ambiguity in language. The Sina Weibo sentiment evaluation (Ling et al., 2020) which adapted a hybrid neural community version for coping with the polysemy phenomena of phrases and subject matter confusion. This approach used the latent semantic relationships in exceptional linguistic contexts. Author used the various window size filters to extract the high stage phrases through multichannel CNN. This proposed version gives higher precision, recall, and F1- rating for weibo sentiment evaluation.

In the article by Obiedat et al. (2022), the researchers utilized hybrid optimization strategies in dealing with skewed sentiment datasets which enhanced the accuracy of the prediction of data in real world contexts of review users. The higher the frequency of individuals visiting social networking sites to interact with companies (restaurants), which increases the frequency of creating reviews on their experiences. Nevertheless, imbalance of data is generally a problem in sentiment analysis and machine learning. The research was conducted using a set

of reviews written by Jordanian restaurants. In order to optimize the weight of the features in the data set, they used PSO and incorporating the SVM into the process. They utilized several high-performing oversampling techniques, including the Synthetic Minority Over-Sampling (SMOTE), SVM-SMOTE, adaptive synthetic sampling (ADASYN), and Borderline-SMOTE. The author focused to balance out the class data set (i.e., negative and positive reviews) using PSO.

Ang et al. (2018) also studied emotion recognition as a mentoring tool in the virtual classroom educational feedback through face recognition. One of the ways is to examine the comprehension of the whole distance learning process amongst students through a facial recognition technique. The model of learning emotion recognition is structured in accordance with three stage architecture: Feature extraction, subset features, and emotion classifier. Virtual classrooms have been created without any compromise in terms of speed or fidelity since the authors came up with a model that was meant to identify emotions through facial expression. Their system has the balance right by employing HAAR cascades to detect eyes and mouth in a fast manner and neural networks to be able to identify emotions accurately. This leads to a system that can capture a broad range of emotions within a short amount of time and with high precision and is meant to enhance interaction and feedback in online classrooms.

Multimodal studies, such as Saxena et al. (2020) established that incorporating multiple modalities drive stronger and robust systems in multimodal emotion recognition. Multimodal modalities are facial expressions, speech alerts, physiological alerts, and textual content semantics. In facial emotion detection methodologies, the SWT put up the highest accuracy of 98.83% with the help of the twin random wooded area classifier. For speech emotion recognition, Biogeography-based optimization using the particle swarm optimization obtained excellent results with 89.47% accuracy on BES dataset, and the Hidden Markov Models had strong generalization performance with 94.41%. Physiological signal-based detection techniques using SVM and coarse set used to achieve an accuracy of 87.15% and 87.02%. Text semantics, as well as facial detection, has exhibited high accuracy at an accuracy of 98.83%. Perhaps, audio-based emotion recognition has significantly progressed in the recent past. BBO aided by particle swarm optimization appeared to be the best approach, and indeed it adequately showed off its muscle in attaining the state-of-the-art results over datasets and modalities.

Following this trajectory, Weiqing Wang et al. (2020) contributed emotion recognition by synthesizing FER algorithms together in online learning platforms in support of providing virtual education. Student's images were captured and classified the emotions by using FER algorithms. This novel paradigm not only applied to virtual classrooms but also augmented online conferences and thus proved to be an all-purpose tool for emotion recognition in real time. A simplified system was built by the researchers by combining an online learning environment with a light CNN-based deep learning architecture that made possible computer-simulated analysis of student affect from facial expressions. The findings are displayed in an intuitive histogram, enabling teachers to immediately sense classroom mood and tailor their teaching style for improved engagement. During experimentation, the system accurately identified 12 faces, with 11 being good, analyzable features. Each recognizable face was analyzed by the CNN model, creating individual emotion tags and a general picture of the emotional state of the class at any instant.

At the same time, advances in the development of sequential models, like RNNs, aim to lessen delays in recognizing dynamic expressive behavior. A Recurrent Neural Network (Baddar et al., 2022) for spontaneous prediction and guided early prediction by sub-sequences through new learning constraints. It is used for facial expression dynamics encoding, which enhances the accuracy of facial expression classification and recognition. Most of the face dynamics-based methods are considered temporally segmented sequences before prediction.

Samadiani et al. (2021) developed the Happy ER-DDF system, an advanced hybrid deep learning framework that accurately detects happiness in natural video settings. The innovative approach combines three key components: (1) an enhanced 3D Inception ResNet architecture that captures both spatial and temporal facial features, (2) an LSTM network that analyzes expression dynamics across video frames, and (3) CNN-based geometric analysis of facial features. This integrated solution demonstrates superior performance, achieving 95.97% accuracy on the AM-FED+ dataset, 94.89% on AFEW, and 91.14% on MELD conversational videos - significantly outperforming traditional techniques like standard CNNs and LBP methods. The system's real-world effectiveness stems from its ability to process spontaneous facial expressions in diverse environments, from casual YouTube videos to video conference calls, by simultaneously evaluating immediate facial characteristics and their temporal evolution while implementing multi-level validation for reliable results.

Building on this point on video-based approaches Rodriguez et al. (2022) Analyzed the VGG facial features using CNN and temporal relation between video frames are detected using RNN+ LSTM hence author demonstrated the detections of human pain expression using various facial features and compared with existing classifier which results the 10% greater accuracy using CK+ dataset. Shojaeilangari et al. (2015) have presented a new non-linear classification approach known as Extreme Sparse Learning (ESL), which delivers high correct classification rates for emotion - even in noisy datasets or

imperfect samples. ESL achieved this by integrating the sparse representation of images (as in a dictionary learning framework), which allows it to perform better than other models or theoretical approaches, in conditions that typically yield low performance. The method cleverly integrated motion of the facial features through the optical flow feature extraction process that carefully tracked features move and morph throughout each video frame. The ESL was compared to several well-known models (SVM, SELM, KELM) on the following four difficult datasets: CK+, ECK+, AVEC, and EmotiW. The results consistently showed ESL's effectiveness in "real-world" problems that routinely produce low classifications and misclassifications in emotion recognition systems that are not robust for noise and variable. Especially relevant is that ESL still maintained high levels of accuracy even in functional bad conditions. Integrating existing approaches for action in this manner, ESL is an advance on previous approaches for use cases related to real world emotion recognition. According to Yao et al. (2021), the current version of EBIR did not examine either the interest area or the hierarchical relationship between the different types of emotion-related information used as stimulus for recovering visual information. Additionally, the current EBIR only looks at the "coarse" conceptual portions (levels of emotional content) instead of the "fine" categories. In the image retrieval process, both polarity level features and emotion level features are treated equally. The proposed method performed a dataset comparison on Flickr and Instagram, ArtPhoto, Abstract paintings (Abstract), and International Affective Picture System (IAPS), against a CNN model for image retrieval and other traditional methods.

## Materials and Methods

### *Experimental Framework and Workflow Design*

Framework for Experiments and Workflow Design We designed our research methodology to evaluate a new deep learning (Fuad et al., 2021) model for categorizing human emotions based on facial images. The process was broken down into three phases: Preprocessing & curating a multi-source image dataset, constructing & training our suggested neural network structure, and lastly, a complete benchmark test versus a panel of established machine learning techniques. This configuration ensured that performance and generalizability of our model were tested equitably and comprehensively.

### *Data Curation and Model Architecture*

Organization of Data and Construction of Models We collected a large and diverse dataset from public sources known for facial expression work, to test and train our models (Choi et al., 2018). This amalgamated dataset brought together images from the Extended Cohn-Kanade (CK+), the Japanese Female Facial Expression (JAFFE), the Berlin Emotional Speech (BES), EMOTIC-W, and the International Affective Picture System (IAP). This combination was extremely important, because it allowed for a wide variety of participants' demographics, lighting environments, and intensities of emotions to be included. We labelled each of the images with one of the seven universal emotions. We generated our classifier using a custom DenseNet-169 architecture but replaced the default ReLU activation function with the SWISH function. This can produce better training dynamics. As this modification could help enhance gradient flow and possibly improve training dynamics during learning, all the images underwent standard preprocessing of resizing and pixel normalization to ensure uniformity in the dataset before entry.

### *Benchmarking and Performance Evaluation*

Performance Evaluation and Benchmarking The main part of our evaluation was a direct performance comparison between our proposed model (SWISH DenseNet-169) and other popular algorithms in the field. We chose a group of models that are well-known for doing well at image classification and regression tasks. These models include Efficient Net, a standard Convolutional Neural Network (CNN), Vision Transformer, a Region-Based RCNN, and a LSTM network set up for image classification (Khan et al., 2019). To avoid bias in performance, we trained and tested each comparator model on the same subset of our composite dataset. The primary metric for this head-to-head comparison was classification accuracy, which easily demonstrated how each model was able to accurately label the fundamental emotional states. The stringent benchmarking process was important to demonstrate how our proposed model outperformed others. While the upper limit of epochs was 30,000, early stopping was incorporated into the model based on the validation loss to stop over fitting. In practice, convergence was reached in 1000 to 1500 epochs. The high upper epoch limit was intentionally chosen to achieve adequate convergence in all experimental conditions while maintaining a low computational cost of the compact model.

### *SWISH DenseNet-169*

The DenseNet-169 architecture is contains multiple layers which compresses with minimum of six convolution layers, twelve max pooling layer, thirty two dense layers and thirty two transition layers and this entire architecture uses the two major activation function called as ReLU and softmax for triggering the threshold values which can be passed to the next levels of layer or output layer (Ling et al., 2020). The image data is received by the input layer; for RGB images, this is usually in the format of 224x224x3. The input goes via

the convolutional layer which can have 64 filters and 7x7 kernel size along with ReLU activation function for reducing the dimension. Next it flows through each dense block, which is made up of tightly connected layers within blocks. The size of feature maps between dense blocks is decreased by transition layers. Eventually, global average pooling is used before categorization after going through every block. Figure 1 dipctes the SWISH DenseNet-169 architecture.

### *SWISH DenseNet-169 Architecture*

Traditionally, the text data take a dominant position in the opinion mining. This has over time proven not to be productive. In text reviews, the detection of emotions is conducted with the help of various algorithms depending on the keywords contained in the information. It is normally applied to analyze the reviews and understand the trends of social media. When making this analysis, verbal forms of communication are not the only types of communication that are taken into account; nonverbal communication should also be taken into account. The problem is that it is imperative that they take into account nonverbal means of communication since the complexity of data increases and the necessity of more accurate findings grows. Moreover, only a few industries can be covered by the traditional technique of opinion mining (Tsao et al., 2010). It has been proposed to use SWISH and Log-softmax in the place of the ReLU and softmax activation functions in order to enhance the accuracy of the multi-class image classification, which is stated in the Eq. 3. To enhance the accuracy of the classification, the DenseNet-169 was modified in the activation function.The proposed SWISH DenseNet-169 Architecture is presented in Figure 1, it provides visual representation of how the different software Components of the System are linked, and how all of the various constraints and how the Components Work together. SWISH activation function can be used in place of ReLU in smooth and non-monotonic operations of deep learning, as described in the Eq. 1 and 2 (Fuad et al., 2021) to perform processes like prediction and classification, namely, feature-based image classification. The role of SWISH is as follows:

$$f(y)= y*\text{sigmoid}^{(\beta*y)} \quad (1)$$

$$\text{sigmoid}(\beta*y)=\frac{1}{1+e^{-(\beta*y)}} \quad (2)$$

Where, β is a scalable & trainable parameter SWISH can be used to improve sample generation in Generative adversarial networks by applying it to both the discriminator and generator models. It works on more complex data patterns and brings the fine gradient flow while stumbling of neurons at inactive state of deep networks. It provides the unbounded outputs like ReLU in addition to smooth flow of data tuning (Obiedat et al., 2022). Log- softmax activation function is used for output triggering in the output layer for multi-class classification problem. It gives back the probability logarithm, which is helpful for numerical stability, particularly in loss functions like NLLLoss:

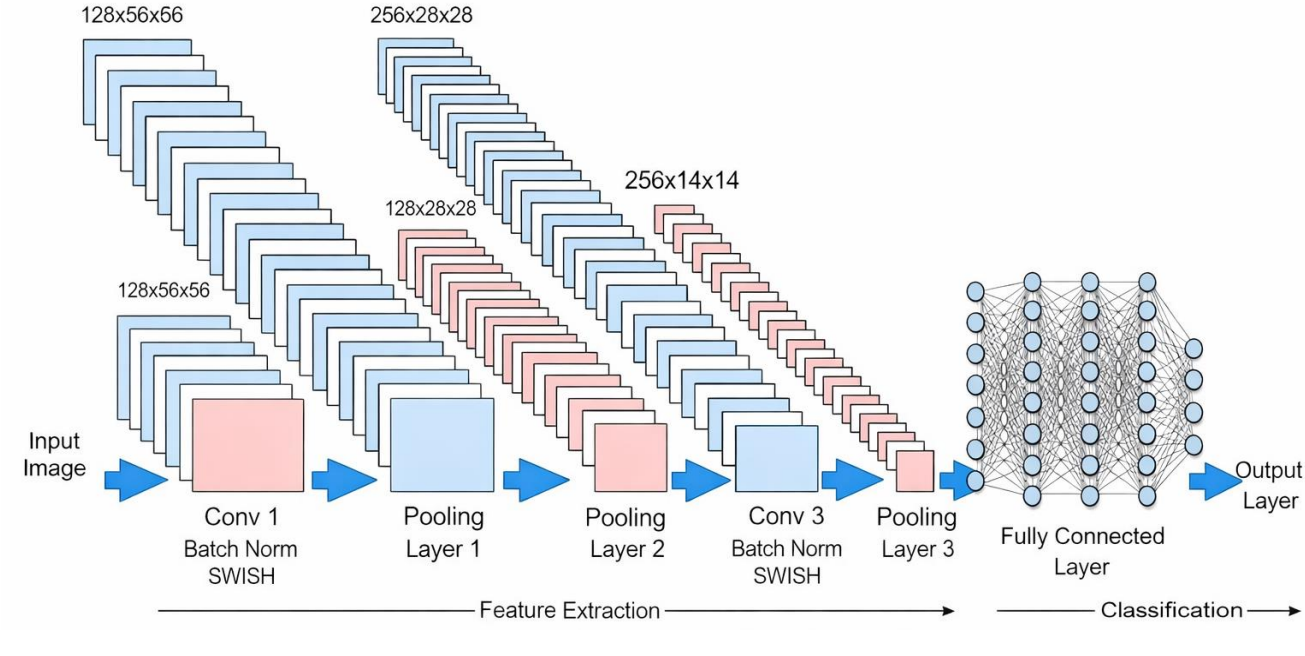


**Fig. 1:** SWISH DenseNet-169 architecture

$$\text{Log-softmax}(y) = \log\left(\frac{e^{y_i}}{\sum_j e^{y_j}}\right) = x_i - \log\left(\sum_j e^{y_j}\right) \quad (3)$$

The prior Figure 2 displayed the SWISH activation function with values of the scaling parameter β (0.5, 1.0, 1.5, and 2.0,-2,-1). The x-axis is the value of the input x and the y-axis is the activation output that corresponds with the input. When β gets larger, the curve becomes progressively steeper, causing transitions around zero to be sharper; smaller values of β produce smoother curves. The SWISH function has negative values that are small; this is important because it alleviates the vanishing gradient problem that occurs with the ReLU function. Figure 6 illustrates the impact of scaling parameter β on SWISH. Each subplot represents a different β value (0.5, 1.0, 1.5, and 2.0). A smaller β (e.g., β = 0.5) produces a smoother function that behaves linearly, while larger β values (e.g., β = 2.0) produce curves that resemble the ReLU activation and cause steeper transitions. Importantly, SWISH does not zero out all negative values, allowing for small negative activations which makes SWISH smoother than the traditional sigmoid activation function and alleviates problems with vanishing gradient.

### *SWISH DenseNet-169 Algorithm*

The Architecture has developed to be highly scalable in Nature to allow for expansion of the Data set and complexity of the Problem domain. The formal Specification of the SWISH DenseNet-169 Algorithm has been implemented in this architecture.

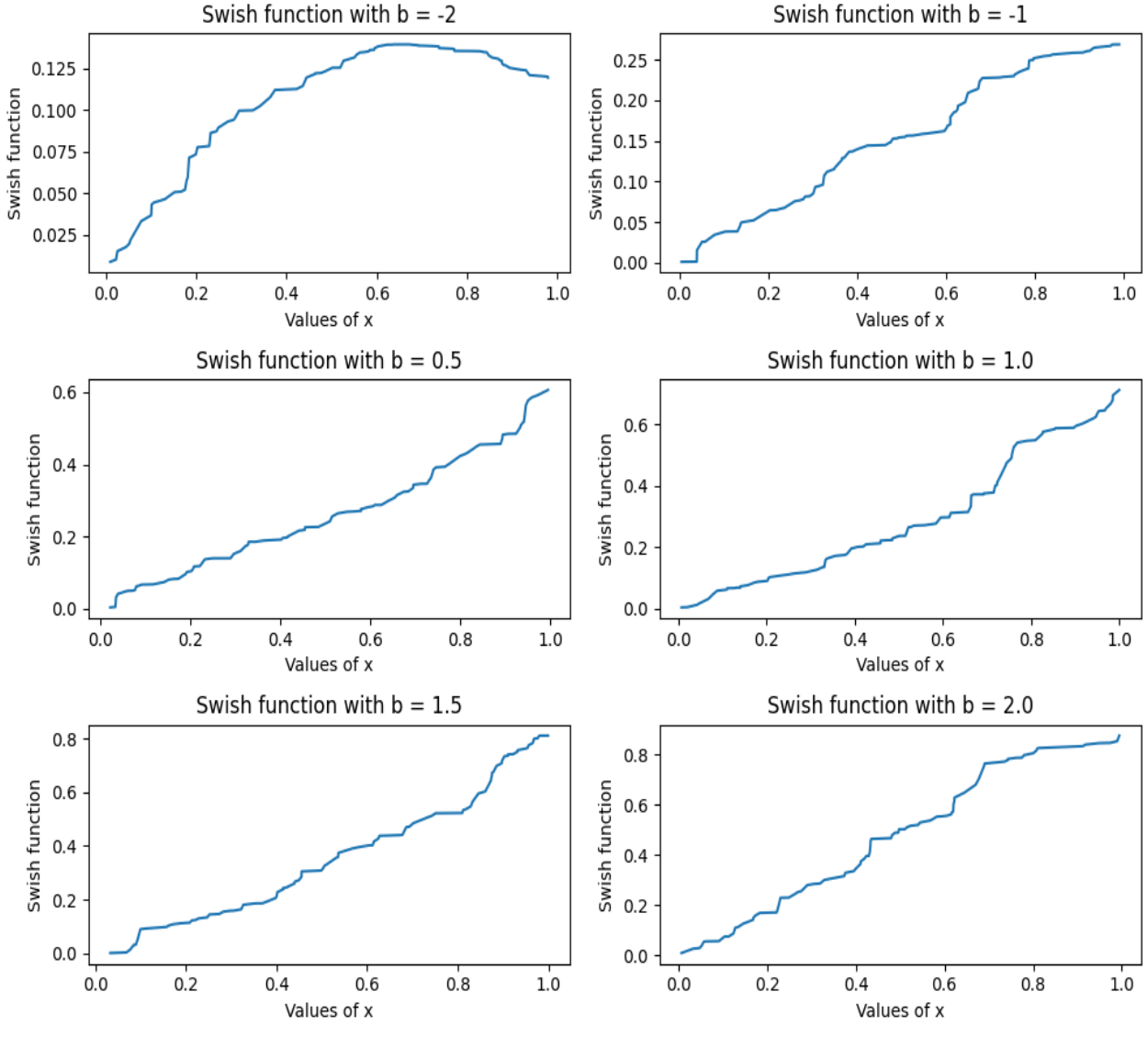


**Fig. 2:** SWISH activation function for different β values

INPUT: Image data X, the number of laye rs L, growth rate k, output classes C

OUTPUT: A probability distribution over C classes

Step 1. Process the input image:
- Step 1.1 input image to single convolutional layer
- Step 1.2 batch Normalization
- Step 1.3 apply SWISH
- Step 1.4 apply max pooling to down-sample image

Step 2. Build first dense block (6 layers):
- For i in range (L-Layer):
  - Step 2.1 normalize input
  - Step 2.2 apply SWISH
  - Step 2.3 create a bottleneck for 1 x 1 convolutional layer
  - Step 2.4 apply drop out
  - Step 2.5 normalize outputs
  - Step 2.6 apply SWISH
  - Step 2.7 apply 3 x 3 convolution to create new features
  - Step 2.8 apply drop out
  - Step 2.9 concatenate features in dense layer
- End For

Step 3. Add transition layer following the first dense block:
- Step 3.1 normalize features
- Step 3.2 apply SWISH
- Step 3.3 down-sample feature count
- Step 3.4 apply drop out
- Step 3.5 apply average pooling

Step 4. Build second dense block (12 layers):
- Step 4.1 repeat process from step 2 but for 12 layers

Step 5. Transition layer before final output:
- Step 5.1 repeat process from step 3

The input is provided in form of an image which is captured either statically or dynamically, and the output is categorised into the categorised emotions. This study applies the image processing layer into converting the colour image into a grey scale image and once this has been converted, the bounded box to surround the face in the image is drawn to locate the face. The first layer is known as the Initialisation Layer and is regarded as the Primary Workflow Layer since it is responsible for creating the Emotional Response Model. The purpose of the Emotional Response Model is to predict what emotions are, and this is done with the Darwin Facial Emotions (Jaiswal et al., 2020), which allows for classifying Emotion into 7 broad categories, which include: Happiness; Sadness; Anger; Fear; Surprise; Disgust; Neutral. The second layer is called the Optimisation Layer. In addition to refining the parameters of the Emotional Response Model, the Optimisation Layer also optimises the processing of the Evaluative Measures as to deliver accurate, reliable evaluations. The third layer is the Deployment Layer. The Deployment Layer represents an actual application of the Emotional Response Model, in which the image captured from the real world is utilized, captured, and fed back into the Emotional Response Model to obtain the predicted emotion. The result of this layer is categorized under the 7 basic emotions. Based on the above suggested algorithm, the input into the network is a collection of image data X, and the output is a probability distribution on C classes. The network is made up of multiple layers with each layer performing a particular operation on the input data, 2x7x7 is the size of the kernel and the stride is 2x2 which is used in the initial layers of convolution after which Batch normalization and Log- softmax activation activation are implemented to the output of this layer.

The max pooling is implemented using the 3x3 pool size and stride of 2x2. Each of the dense blocks that make up a block building is fed into the algorithm following the first convolution layer then the loop that forms L dense blocks. each dense block itself, the output feature maps of each of the L following layers are concatenated together and fed to the final layers as input. Each dense block contained the same standard quantity; thus, the construction of each block followed a standard pattern: Each block had Batch Normalization at the beginning, followed by SWISH, and followed by 1x1 Convolutional Layer with 4k number of filters. Finally, there is a Classification Module with a fully connected layer after the Global Average Pooling and a Log-softmax activation function to generate the final probability distribution of C number of classes.

The DenseNet architecture is designed to densely connect the layers of a convolutional neural network in a feed-forward fashion, enabling more efficient and effective learning by propagating information more directly across the layers and encouraging feature reuse. SWISH gives the smooth, stable gradient flow in functions like GELU and Mish, which is crucial for training a deep network effectively. But it also has that non-monotonic 'hump' for negative values, similar to Mish, which helps the model capture the subtle, complex patterns in facial expressions. The clincher for us was SWISH's trainable beta parameter. This lets the network subtly adjust the function's behaviour for our specific task, adding a layer of customization that other functions lack and it does all this without being overly complex to compute, which keeps our training times manageable. This architecture has shown state-of-the-art performance on several benchmark image classification datasets. The proposed model contains the four major modules such as image processing, initialisation of prediction, optimisation and model deployment.

### *Image Pre-Processing*

In order to have strong and robust emotion prediction, pre-processing pipeline of the images has been implemented extensively before the models are trained. Face detection, alignment, cropping, resizing, grayscale conversion, normalization and data augmentation are used to constitute the pipeline as explained below.

### *Face Detection and Alignment*

The Haar Cascade frontal face detector (Ghali et al., 2018) was employed to identify face parts because it is highly used in real-time face detection as it is computationally efficient. The detector was set with the scale factor of 1.1, lowest neighbours of 5 and a minimum size of face of 30 x 30 pixels as a compromise between the accuracy of detection and false positives. After a face was recognized, the bounding box was created around the facial area, and all the further processing was limited to the region of interest. This step of cropping helps to remove the background noise and only discriminative features of the face were considered in categorizing emotions (Yang et al., 2018). Basic face alignment which is based on detected facial landmarks namely the eye locations was done to minimize variations due to head pose and in-plane rotation. The identified face part was oriented and positioned in a way to make the eyes horizontally aligned. It is done to increase spatial consistency in samples and make the model learn emotion-specific patterns. Figure 3 classifies these emotions into discrete categorical groups, providing the foundational labels for the detection pipeline. Figure 4 illustrates how this feature detection works in reality, converting raw pixels to quantifiable emotional indicators.


**Fig. 3:** Classification of Emotions

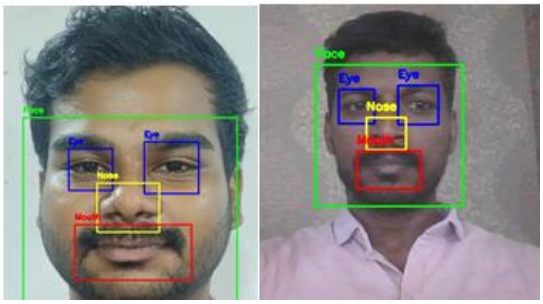


**Fig. 4:** Bounding Box feature extraction

### *Grayscale Conversion and Normalization*

The RGB-to-grayscale conversion (luminosity approach) was used to convert the cropped and aligned face images, which approximates the human visual perception better than a simple averaging approach does. The grayscale value, I, was calculated as a linear sum of the RGB values: I = 0.299R+ 0.587G +0.114B. This weighting gives more weight to the green channel as the human eye is more sensitive to green wavelengths, and it retains fine details of changes in intensity that are essential in reading facial expressions (Saxena et al., 2020). All grayscale face images were scaled to a constant spatial resolution of 48 x 48 pixels to have the same dimensions of input to the SWISH-activated DenseNet-169 model. The pixel intensities were then brought to the range [0, 1]/255. This stabilizes gradient updates during training and models convergence.

### *Data Augmentation*

To enhance generalization and reduce overfitting, the training set was augmented with data. Horizontal flipping, small random rotation and slight scaling alteration were considered as augmentation operations but the semantic meaning of facial expressions remained the same (Jaiswal et al., 2020). There was no augmentation of the validation set or test set to be fair in performance assessment.

### *Emotion Prediction Using SWISH DenseNet-169*

ResNets' architectural design approach is sequential and cumulative. Each layer receives input directly from its preceding layer. DenseNets, however, establish direct links between their layers enabling each layer to be connected to every upstream layer. This creates a dense interconnection graph that accumulates features through the depth of the entire network. This approach specifically addresses the greater concern of the vanishing gradient problem in deeper networks, where the essential information can get "lost in transit" from deeper layers of the deep network. DenseNets preserve these close connections so that the early layer information can be carried all the way through to the final output layer. This design is its simplicity: It uses simple connectivity rules for a dramatic effect on memory capacity throughout all of the depth of the network (Ling et al., 2020). This creates a stark visual contrast compared to ResNet's more linear flow of information, which is partly DenseNets will typically develop more accuracy in their deep network implementations.

It builds a dense network with l(l+1)/2 connections for a given l, where every layer includes input from all of the previous layers. The optimization trajectory is governed by a critical hyper parameter named as learning rate. In this study, key hyperparameters including the model architecture, learning rate, epoch count, and random seeds dictate the training process. The seeds, in particular, initialize the model's

parameters to a reproducible state. All configured values for these hyperparameters are documented in Table 1. Model of this emotion prediction based on SWISH-activated DenseNet-169 was introduced and trained in controlled and repeatable conditions. All experiments were run with a constant random seed of 12 to bring consistency between runs. All the images of the faces were trained at 48 x 48 pixels and normalized before being trained. The entire data is of 6,078 images with seven emotion categories, i.e., anger, disgust, fear, happiness, neutral, sad and surprise. The data were split into training, validation, and test sets:

Training set: 70% of images
Validation set: 15% of images
Test set: 15% of images

This same division was carried out throughout all the experiments. Adam optimizer having an initial learning rate of 0.01 was used to train the network. Learning-rate scheduling was used in the fine-tuning process, this meant that the learning rate was gradually decreased as the number of epochs exceeded 25,000 to stabilize the convergence. The maximum number of epochs that the model was trained was 30,000 and early stopping was implemented with a patience criterion of 3 epochs in terms of validation loss to avoid overfitting. The training was done with a batch size of 64. To enhance generalization, dropout regularization was used in the fully connected layers and L2 weight decay was used in the optimization. The last classification layer puts out probabilities of 7 emotion classes through a SWISH activation function. Figures 4 and 5 illustrates feature detection and quantifiable emotional indication using SWISH DenseNet-169 . Optimization was done using the categorical cross-entropy loss. Each reported result came out of one deterministic run using a fixed seed, meaning it is reproducible. Training was monitored on validation accuracy and validation loss, and the final model was chosen based on the optimal validation before evaluating it on the held-out test set.

## *Optimization*

There are two parts to reach higher accurate models: Fine-tuning and training. The model learns optimal weight and bias values based on the labelled data during the training phase (this is an extremely simple step in Machine Learning). The model produced is considered to have potential and can be tested, validated, and used in future (Wang et al., 2020). The model's performance relies heavily on training error reduction, which is contributed by quality training data and the choice of the algorithm for training model parameterization.

**Table 1:** Training hyperparameters of the SWISH DenseNet-169 model

| Parameter | Value |
|---|---|
| Optimizer | Adam |
| Initial Learning Rate | 0.01 |
| Learning-Rate Schedule | Reduced after 25,000 epochs |
| Batch Size | 64 |
| Total Epochs | 30,000 |
| Early Stopping | Patience = 3 |
| Weight Decay | L2 regularization |
| Dropout | Applied in dense layers |
| Number of Runs | 1 |
| Random Seed | 12 |
| Train/Val/Test Split | 4,254 / 912 / 912 |
| Image Size | 48 × 48 |
| Number of Classes | 7 |

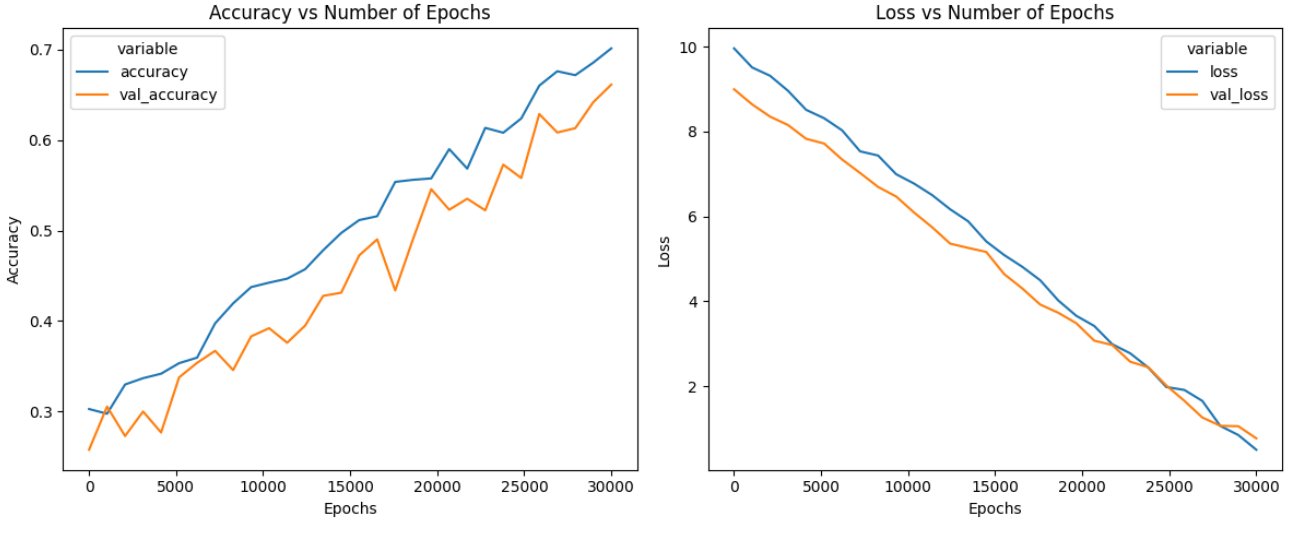


**Fig. 5 (a-b):** Training and validation performance of the SWISH DenseNet-169 model

Evaluation data is divided into three portions: 70% training, 15% validating, and 15% testing. A device offers developers the advantage of effectively configuring and testing their model on an unseen dataset during the creation of machine learning models and tuning is also another way of enhancing the performance of the model after training. In many situations, deep learning is a method of fine tuning previously learned knowledge to assist in solving other, more general problems. Ultimately, this means that deep learning allows us to "fine tune" a model based on the results that were achieved through experience with a prior task. Ultimately, deep learning represents an important step towards improving the accuracy of predictions (Yao et al., 2021) and is also a critical component for refining the model.

The plots above show the model's performance over a training and validation phase of 30000 epochs. Figure 5 (a) Accuracy vs epochs: For both training and validation, accuracy began to increase immediately and then quickly plateaued after about 15000 epochs, indicating the model is converging. Figure 5 (b) Loss vs epochs: The same rapid decline in early epochs followed by rapid flattening of loss plots for both training and validation, indicating that the model is learning well with very little over fitting. The accuracy for both training and validation improved tremendously, especially for the first 10000 epochs, at which point the curve begins to flatten and stabilize around an accuracy of 0.65, indicating that training for longer than approximately 15000epochs would contribute very little actual improvement. The right plot shows the loss curves where both training and validation loss also decreased dramatically through early epochs, before flattening out and converging to a minimum loss. Both plots indicate the training and validation curves track closely with one another, which indicates the model has generalized well and is likely applicable to unseen data without a substantive concern for overfitting.

The images in Figure 6(a&b) showcase different variations of the SWISH activation, based on the value of the variable `β` (positive and negative). The left-hand side (plot 6a) shows the effect of increasing a positive value of `β` on an otherwise flip curve. When `β` increases, the slope of the curve also increases (superlinear), as shown by plot 6a, and approaches the shape of ReLU (Rectified Linear Unit) as the value of `β` continues to increase. The right-hand side (plot 6b) shows both positive and negative values for `β`, using values of `β` from the original data set. Negative values of `β` flip the curve such that it is decreasing rather than increasing, resulting in negative activation flows through both negative and positive regions of the curve. On the left-hand side, positive `β` values will produce a bounded SWISH activation function that is smooth and will retain its non-linear shape and behaviour as well as being non-monotonic. As the size of the parameter β decreases, as seen for example in case of `β` = 0.5, the shape of the curve will become flatter, whereas larger `β` values will result in sharper transitions (e.g., `β` = 2.0) and will resemble the overall shape of a standard ReLU activation function but will allow for some small negative activations. The importance of these transitions is that they help to mitigate the issue of the vanishing gradient with the training of deep networks as it allows information to remain in the negative activation area throughout training.

The Figure 6(b) uses a negative β (β=−1.0 or -2.0) to flip the shape of the SWISH activation function. This inversion of the curve yields a decreasing SWISH curve that suppresses positive inputs and magnifies negative inputs. Again, having a negative β value being implemented is rarely used in practice but demonstrates the immense flexibility of the SWISH formulation. The above plots demonstrates how β is akin to a scaling parameter.

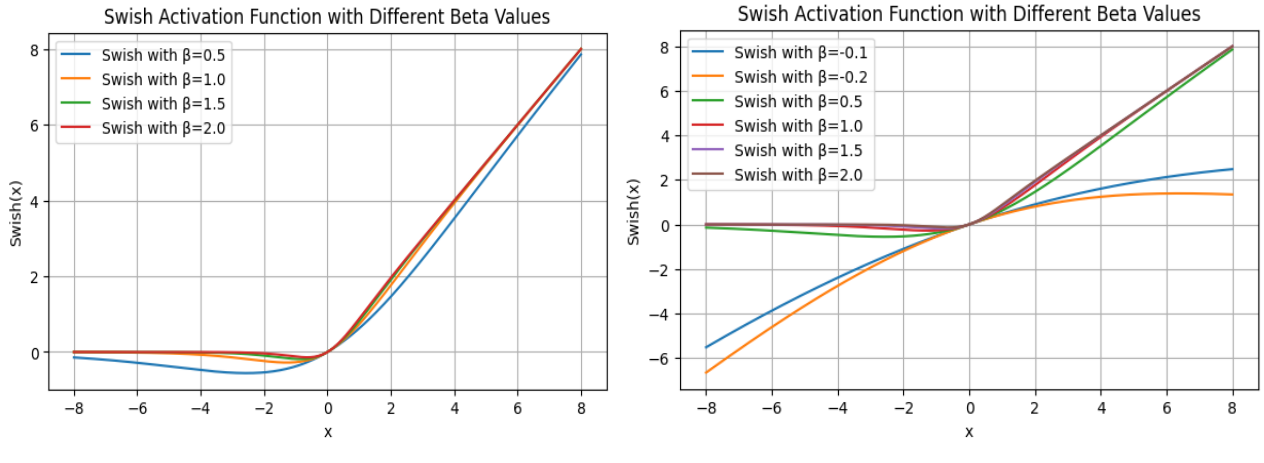


**Fig. 6 (a-b):** SWISH activation function with different negative β values

A smooth, non-linear function called the SWISH activation function has been suggested as a substitute for well-known activation functions like ReLU (Baddar et al., 2022). Because it is non-monotonic, the function may be both decreasing and rising for certain input ranges. The standard SWISH activation function consistently employs a beta value of 1.in this case, the function becomes:

$$\text{SWISH}(y)=\frac{y}{1+e^{-y}}=x.\sigma(y) \quad (4)$$

Where σ(*y*) is the sigmoid function.

This is the conventional SWISH; for positive values, it operates like ReLU, but for negative inputs, it is smoother and accepts small negative values. Positive values of β, SWISH tends to get sharper and then non-linearity intensifies to attain more closely to the ReLU activation function as beta (β) rises. SWISH, in contrast to ReLU, nevertheless permits negative values, which in some circumstances could facilitate gradient flow during back propagation. The higher values beta (β) forms the hard threshold function (Samadiani et al., 2021), which results in a stronger transition around x = 0. The SWISH function's structure changes dramatically when beta (β) is negative. When inputs are positive, the output is mostly negative; when inputs are negative, the output is positive as depicted in the Figure 6(b). SWISH's ability to alter (β)beta and tune for various neural network designs can result in a reversal of behaviour, where the function favours passing negative values while suppressing positive ones. SWISH reverses the roles of positive and negative inputs, which is advantageous in some topologies where it is preferable to have the opposite activation response. The model is capable of learning an ideal activation function due to (β) beta's flexibility.

## *Forecasting Events from Visual Data*

This research proposes a computational process to forecast emergent collective human events like protests and large public gatherings by analyzing unstructured video and image data. The method comprises three stages which is depicted in the Figure 7 such as affective computation, feature abstraction, and predictive modeling.

This architecture takes both emotional and temporal aspects of the user into account to develop an accurate prediction of the user's activity/event. Each portion of the multi-pipe architecture was created as follows: The raw image data being sent to the Emotion Recognition System (Emote). The Emotion Recognition Module (ERM), which has a SWISH activated DenseNet-169 backbone is then used to extract high-level emotion representations with dense feature re-use and nonlinear activation (SWISH) for robust recognition across a wide range of emotional states. Emotion features are then clustered with engineered features containing context or other data relevant to the event prediction task.

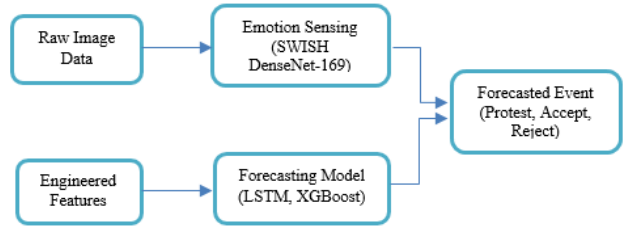


**Fig. 7:** Workflow of the Sentiment Analysis and Event Prediction Model

These combined features are used to create a predictive model using LSTM to model temporal dependencies and XGBoost for the best learning of decision boundaries. The prediction of the user's event will yield either an Accept, Reject or Protest outcome according to the emotional and temporal analyses. This hybrid approach to event forecasting addresses the disconnect between identifying emotions from visual stimuli and creating a sequential model of decision making based on these emotions, thereby allowing for better predictive performance in real-world human event forecasting.

## *Affective Computation from Facial Expressions*

In the first stage, we convert raw pixel data from video data as to a quantitative form of collective affect. We use a facial expression analysis module using SWISH activated DenseNet-169 architecture (Ghali et al., 2018). Due to its dense connectivity, this model is particularly suited for computationally intensive tasks, while enabling reuse of features across layers, both for individual facial landmarks and in their global articulation. To promote better learning behaviour and limit gradient vanishing, the model uses the swish activation function. It was chosen because it provides a more normalised gradual increase in transition for modelling nonlinear relationships than other functions with the steep mapping such as the common Relu function. The output it generates is a continuous multi-class time-series representation of the spectator population's emotional states over time (anger, joy, fear, sadness, surprise).

## *Temporal Dynamic Features*

Predictions based on direct emotional probabilities are frequently too erratic to be reliable. In this regard, we build a set of higher-order features that represent the temporal and statistical properties of the emotional time-series (Rodriguez et al., 2022). This level of abstraction is necessary for this context and to better distinguish behavioural precursors from context noise. The process involves constructing a compact and informative feature space for prediction through the extraction of three features of intensity trends (subset of long-range directionality of emotional change), volatility index (variation in group engagement), and anomalous peaks (rapid quick change compared to any event).

*Predictive Modeling of Event Likelihood*

The final stage of the modelling process involves training predictive models on the engineered features. There are advantages to both modelling types, a Long Short-Term Memory (LSTM) network is used to capture sequential dependencies in the data (Shojaeilangari et al., 2015), and emotional trajectories changes over time scales. The second model deployed is an XGBoost model, which is robust with structured tabular data and can also model complicated non-linear interactions among features. Selecting a model or combining both will depend on the forecasting horizon, and the characteristics of the event being modelled. The pipeline produces a probabilistic estimate of a future group occurrence. It provides an estimate of risk, or safety, in space and time. The primary use case is operational early-warning systems to give decision-makers crucial time for taking preventive action, providing more analysis, or undertaking non-escalatory options.

## Results Analysis and Interpretation

The proposed model was tested on a composite test set consisting of 6,078 images that were gathered in five popular affective datasets, i.e., CK+, JAFFE, BES, EMOTI-W and IAPS. All these datasets represent seven emotional categories, namely, neutral, happy, surprise, anger, disgust, afraid (fear), and sad. The application of various datasets provides different identities of the subjects, different acquisition conditions, and different emotional expressions thus minimizing the chances of dataset-specific or class-dominance bias. The existence of variation in the frequency of emotion of the individual datasets, although, the combination of multiple sources of sample leads to a balanced proportion of the overall class distribution of the entire seven emotions. This combined assessment plan will reduce the effects of the imbalance of the classes on the presented mean accuracy and provide a more trusted measure of the ability of the model to generalize.

Fig. 8 displays the confusion of the proposed model, which shows the classification results of the seven emotive categories. The rows in the prediction reflect the ground-truth emotion classes (Jaiswal et al., 2020), and the columns display the predicted ones. Value high on the diagonal means the correct classifications and value low on the diagonal means the misclassification percentages. The model has high level of discriminative ability which is indicated by the high levels of diagonal values in most categories of emotion. The surprise emotion is identified with an objective accuracy of 100% indicating that the peculiarities of its face and context are well represented by the supported representation. Likewise, the recognition rates of happy (95.16), sad (95.56) and fearful (94.22) emotions are high and therefore the learned features are strong in both low- and high-arousal affective conditions. The neutral category has an accuracy of 88.17, and there is a small amount of confusion especially with happy (2.62%), disgust (2.29%), and sad (3.07%). This phenomenon is explained by the fact that there are delicate visual resemblances between a neutral expression and low-intensity emotional states, and this difficulty is usually found in the emotion recognition literature.

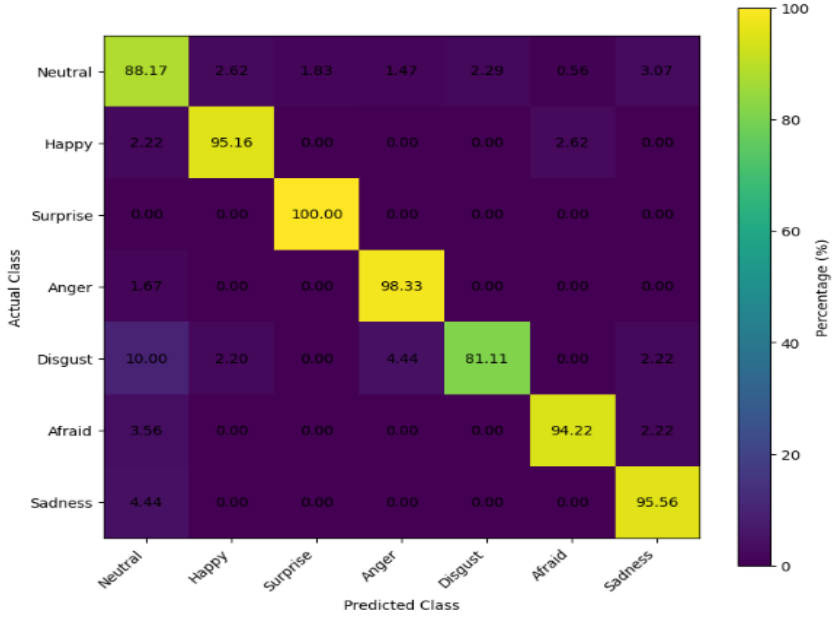


**Fig. 8:** Confusion matrix of the proposed model (actual vs. predicted emotion classes)

The dataset to be used in performance analysis comprises of all emotional images that are divided into seven emotional categories. The distribution by the class can be summed up in Table 2 with the ratio of samples in each of the emotions varying between 13.85% (surprise) and 15.01% (neutral). Such a relatively homogeneous distribution means that no strong imbalance in classes is present, and the reported performance measures will not be biased towards a particular prevalent emotional group.

### *Comparison of Base Line Models*

The Figure 9 illustrates a donut chart that is used to summarize the performance parameters of the SWISH DenseNet-169 model when used in emotion recognition. Every part of the chart is a major evaluation measure that gives a detailed perspective of how effective the classification of the model would be based on the dataset considered.The overall accuracy of the model is 91.34, which implies that the number of cases of correctly classified emotional instances is high. The accuracy of 93.71% indicates the ability of the model to make very few false positive predictions and its recall (sensitivity) of 90.47% indicates the ability of the model to make correct predictions of the actual classes of emotions. Also, the specificity of 93.31% shows that this model can distinguish very specific cases of rejecting irrelevant categories of emotional categories. The balance between precision and recall of 92.06% of the F1-score further attests to a good and consistent performance. Conversely, false negative rate of 9.53% is also not very high meaning that there has been very little misclassification of genuine emotional states

According to Figure 10, this 3D surface plot is associated with the mean accuracy in predicting the error of emotion recognition of various deep learning models, i.e., CK +, JAFFE, BES, EMOTI-W and IAPS. The datasets are projected on the depth axis (horizontally) and the models of classification (CNN, Vision Transformers, RCNN, LSTM, and SWISH DenseNet-169) are projected on the horizontal axis, respectively. The vertical axis depicts the average of prediction accuracy of recognition and average of the seven categories of emotion. Surface colour code will cut across the ranges of accuracy and this will give the intuitive perceptual visualization of the difference in the performance of the two models and datasets.

**Table 2:** Distribution of Images across Emotion Classes

| Emotion | No. of Images | Percentage (%) |
|---|---|---|
| Neutral | 912 | 15.01 |
| Happy | 876 | 14.42 |
| Surprise | 842 | 13.85 |
| Anger | 851 | 14 |
| Disgust | 863 | 14.2 |
| Fear (Afraid) | 846 | 13.92 |
| Sad | 888 | 14.6 |
| Total | 6078 | 100 |

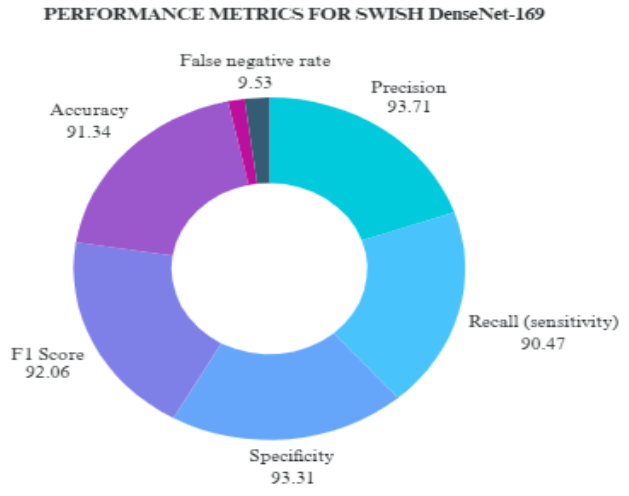


**Fig. 9:** SWISH DenseNet-169 Performance metrics

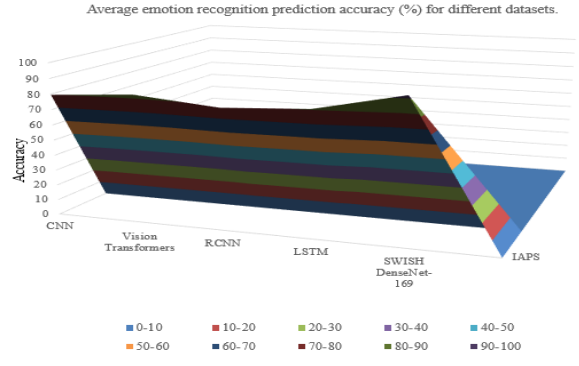


**Fig. 10:** 3D surface plot of average emotion recognition prediction (%) across models and datasets

As can be seen, the topmost layer of each of the datasets was always in the proposed SWISH DenseNet-169 and therefore, it had superior and more reliable recognition performance than the designs of the baseline. Conversely, the former CNN and RCNN models are relatively low-accurate on in-the-wild data sets, like EMOTI-W.

The graph stated in Figure 11 depicts the prediction accuracy of 7 human emotions from various methods. The above-mentioned data sets contain real-time human facial images. The Existing system for Facial Emotion Recognition (Hazar et al., 2013) is solely based on geometrics characters of human facial features. The proposed System uses visual data to predict emotions. While using visual data for emotion detection, the accuracy of the system lies around 91.34%. But the efficiency of the model can be increased further by tweaking the model according to data set selection. While using text data as an input, the data input can be plagiarized but the face cannot hide the emotion. Even though the accuracy is low, the proposed system has laid a foundation for further developments.

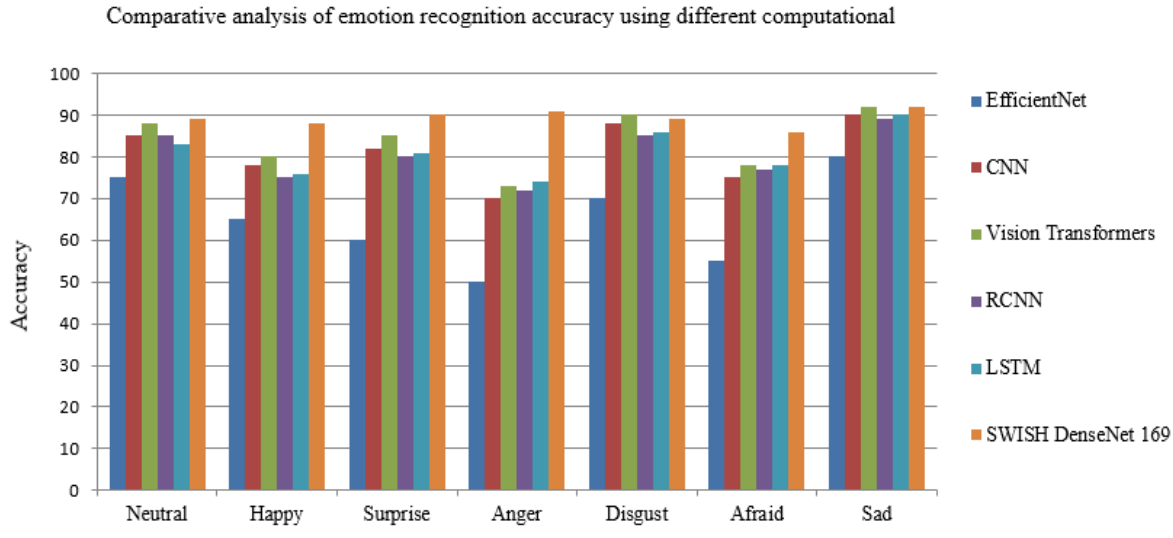


**Fig. 11:** Accuracy comparison of different methods across seven emotional categories

Since the proposed system is a framework, it can be tuned in a way such that it can be combined with existing systems. By combining these two systems we can predict emotions based on both verbal and non-verbal forms of data. This combinational approach can increase the accuracy to a greater extent.

### *Ablation Study*

To study the effects of individual design choices on the proposed emotion prediction framework, ablation experiments were conducted. All ablation tests used the parameters such as backbone (DenseNet-169), training set, processing pipeline, optimizer, learning rate, batch size, epochs, and training-validation-testing configuration. Identical to those in the main experiment, with each component changed separately so that fair comparisons could occur. The initial Ablation test measures the effects of the activation function by substituting the SWISH activation with the standard ReLU, although the network structure remains the same. The findings show that the SWISH-enabled model is always more effective than the ReLU-enabled deep net in any of the assessment measures. One possible reason this is enhanced performance is due to smooth and non-monotonic behavior of SWISH, which is able to provide a better gradient flow and more learning of features in deeper layers. Conversely, ReLU can have dead neurons problems, which restrict its expression ability when it comes to fine discrimination of facial emotions.

### *Output Layer Effect: Log-Softmax vs. Softmax With Cross-Entropy Effect Loss Function*

The second ablation experiment looks into the ultimate decision of output formulation probability by comparing Log-Softmax with Softmax with categorical cross-entropy loss. As experimentally demonstrated, the Log-Softmax set up exhibits slightly superior stability in classification and convergence. The main reason is that this has been improved through better numerical stability in optimization, especially when the emotion classes are very close. However, the cross-entropy version of Softmax also shows that the proposed framework is resistant to the changes in the formulation of the output layer. The third ablation test measures the impact of the input representation by training the model on grayscale images and RGB images respectively. Although RGB inputs do not loss the extra information of color, the configuration based on the grayscale can produce the performance that is similar or even better. This finding indicates that structural facial features of luminance including muscle movements and changes in texture are more predominant in the recognition of emotion, more than chromatic information. Moreover, grayscale inputs are less complex and use less memory, so they are also better adapted to real-time and resource-constrained software.

### *Ethical Considerations and Deployment Constraints*

Though the presented SWISH-enabled DenseNet-169 model proves to be effective, there is a risk of possible fairness weaknesses because it is biased by the dataset. The sample emotion datasets employed might not be sufficiently representative of demographic diversity in terms of age, ethnicity and cultural backgrounds and this may result in uneven performance in practice. Moreover, facial emotion recognition is a sensitive visual information, which creates privacy concerns in terms of data collection, storage, and implementation. Although this research is based on solely ethically obtained publicly available data, practical use must make sure that there are

informed consent, data-handling security, and adherence to the data-protection laws. The use of emotion recognition technologies can bring ethical issues in situations involving intensive or massive surveillance. The suggested solution will have a purely research oriented and helps to society, e.g. human-computer interaction, event forecasting, and must be implemented with openness and reasonable human control. Moreover, the existing system is based solely on an analysis of still photographs. Therefore, it does not consider the changes that occur over time or across different ways of connecting with others that could result from long-term implementations of these systems. Henceforth, future studies should focus on exploring new ways to integrate different types of information as well as how best to create policies governing the responsible application of these systems.

## Conclusion

This study presents a novel method DenseNet-169 which is applied to recognize facial emotions through opinion mining in non-verbal visual information. The input of text or audio was not used, only image-based expressions are used in to the model (Ghali et al., 2018). This research uses a database containing a total of 6078 images assigned to the following primary emotions: anger, disgust, fear, happiness, neutrality, sadness, and surprise. The image database was created combining and organizing the image data of five well-known benchmark databases, including the Extended Cohn-Kanade (CK+), Japanese Female Facial Expression (JAFFE), Berlin Emotional Speech (BES), EMOTIC-W, and the International Affective Picture System (IAPS). The framework presented in this paper was created to classify all seven emotions in the 6078 images in the combined image data set. Before the model training, we carried out a number of image processing techniques to emphasize main facial features. The data was divided in 70-30, where 70 percent was used in training, 15 percent was used in validation and 15 percent to test. SWISH enabled DenseNet-169 was chosen to be used in emotion prediction wherein it was also human action being predicted. The offered model was able to reach a mean prediction accuracy of 91.34% of the examined dataset with seven emotional categories after optimization, which is superior to the existing state-of-the-art methods, such as EfficientNet, CNN, Vision Transformers, and RCNN, LSTM.

### *Future Enhancements*

This model has the potential to perform more effective by adding the hardware and IOT devices. It can also be combined with existing approaches such as fraud detection, product recommendation and Google lens applications to increase the performance much better. Since the proposed model is a framework, it can be fine-tuned according to different applications and different input data. The accuracy of the model is proportional to the data provided and the application it is being used on. SWISH DenseNet-169 is focused on majority of emotion-related tasks but slightly lower on politeness classifications, which we intend to improve in the future.This framework can be incorporated with frontend to develop a complete application for easy use.

## Acknowledgment

The authors would like to state that no support from any organization was received for the preparation or submission of this manuscript.

## Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Author's Contributions

All authors contributed equally to the conception, design, analysis, and preparation of the manuscript.

## Ethics

This manuscript presents original work that has not been published elsewhere. The corresponding author confirms that the study does not involve any ethical issues requiring approval.

### *Conflict of Interest*

The authors declare that there are no conflicts of interest related to the content of this article.

## References

Baddar, W. J., Lee, S., & Ro, Y. M. (2022). On-the-Fly Facial Expression Prediction Using LSTM Encoded Appearance-Suppressed Dynamics. *IEEE Transactions on Affective Computing*, *13*(1), 159–174. https://doi.org/10.1109/taffc.2019.2957465

Cha, H.-S., Choi, S.-J., & Im, C.-H. (2020). Real-Time Recognition of Facial Expressions Using Facial Electromyograms Recorded Around the Eyes for Social Virtual Reality Applications. *IEEE Access*, *8*, 62065–62075. https://doi.org/10.1109/access.2020.2983608

Choi, S.-I., Lee, S.-S., Choi, S. T., & Shin, W.-Y. (2018). Face Recognition Using Composite Features Based on Discriminant Analysis. *IEEE Access*, *6*, 13663–13670. https://doi.org/10.1109/access.2018.2812725

Fuad, Md. T. H., Fime, A. A., Sikder, D., Iftee, Md. A. R., Rabbi, J., Al-Rakhami, M. S., Gumaei, A., Sen, O., Fuad, M., & Islam, Md. N. (2021). Recent Advances in Deep Learning Techniques for Face Recognition. *IEEE Access*, *9*, 99112–99142. https://doi.org/10.1109/access.2021.3096136

Ghali, A. L. I., & Kurdy, M.-B. (2018). Emotion recognition using facial expression analysis. *Journal of Theoretical and Applied Information Technology*, *96*(19), 6117–6129.

Hazar, M., Nesrine, F., Mohamed, H., & Ben-Abdallah, H. (2013). Data Mining-based Facial Expressions Recognition System. *Frontiers in Artificial Intelligence and Applications*, 185–194. https://doi.org/10.3233/978-1-61499-330-8-185

Jaiswal, A., Krishnama Raju, A., & Deb, S. (2020). Facial Emotion Detection Using Deep Learning. *2020 International Conference for Emerging Technology (INCET)*, 1–5. https://doi.org/10.1109/incet49848.2020.9154121

Khan, M. Z., Harous, S., Hassan, S. U., Ghani Khan, M. U., Iqbal, R., & Mumtaz, S. (2019). Deep Unified Model for Face Recognition Based on Convolution Neural Network and Edge Computing. *IEEE Access*, *7*, 72622–72633. https://doi.org/10.1109/access.2019.2918275

Ling, M., Chen, Q., Sun, Q., & Jia, Y. (2020). Hybrid Neural Network for Sina Weibo Sentiment Analysis. *IEEE Transactions on Computational Social Systems*, *7*(4), 983–990. https://doi.org/10.1109/tcss.2020.2998092

Obiedat, R., Qaddoura, R., Al-Zoubi, A. M., Al-Qaisi, L., Harfoushi, O., Alrefai, M., & Faris, H. (2022). Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution. *IEEE Access*, *10*, 22260–22273. https://doi.org/10.1109/access.2022.3149482

Rodrigues, H. R., Gaspar, M. A., & Sassi, R. J. (2020). Opinion mining framework applied to social networks data for small and medium enterprises. *Revista Gestão & Tecnologia*, *20*(3), 167–192. https://doi.org/10.20397/2177-6652/2020.v20i3.1887

Rodriguez, P., Cucurull, G., Gonzalez, J., Gonfaus, J. M., Nasrollahi, K., Moeslund, T. B., & Roca, F. X. (2022). Deep Pain: Exploiting Long Short-Term Memory Networks for Facial Expression Classification. *IEEE Transactions on Cybernetics*, *52*(5), 3314–3324. https://doi.org/10.1109/tcyb.2017.2662199

Samadiani, N., Huang, G., Hu, Y., & Li, X. (2021). Happy Emotion Recognition from Unconstrained Videos Using 3D Hybrid Deep Features. *IEEE Access*, *9*, 35524–35538. https://doi.org/10.1109/access.2021.3061744

Saxena, A., Khanna, A., & Gupta, D. (2020). Emotion Recognition and Detection Methods: A Comprehensive Survey. *Journal of Artificial Intelligence and Systems*, *2*(1), 53–79. https://doi.org/10.33969/ais.2020.21005

Shojaeilangari, S., Yau, W.-Y., Nandakumar, K., Li, J., & Teoh, E. K. (2015). Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning. *IEEE Transactions on Image Processing*, *24*(7), 2140–2152. https://doi.org/10.1109/tip.2015.2416634

Tsao, W.-K., Lee, A. J. T., Liu, Y.-H., Chang, T.-W., & Lin, H.-H. (2010). A data mining approach to face detection. In *Pattern Recognition* (Vol. 43, Issue 3, pp. 1039–1049). https://doi.org/10.1016/j.patcog.2009.09.005

Wang, S., Shuai, H., Zhu, L., & Liu, Q. (2024). Expression Complementary Disentanglement Network for Facial Expression Recognition. *Chinese Journal of Electronics*, *33*(3), 742–752. https://doi.org/10.23919/cje.2022.00.351

Wang, W., Xu, K., Niu, H., & Miao, X. (2020). Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation. *Complexity*, *2020*, 1–9. https://doi.org/10.1155/2020/4065207

Yang, D., Alsadoon, A., Prasad, P. W. C., Singh, A. K., & Elchouemi, A. (2018). An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment. *Procedia Computer Science*, *125*, 2–10. https://doi.org/10.1016/j.procs.2017.12.003

Yao, X., Zhao, S., Lai, Y.-K., She, D., Liang, J., & Yang, J. (2021). APSE: Attention-Aware Polarity-Sensitive Embedding for Emotion-Based Image Retrieval. *IEEE Transactions on Multimedia*, *23*, 4469–4482. https://doi.org/10.1109/tmm.2020.3042664