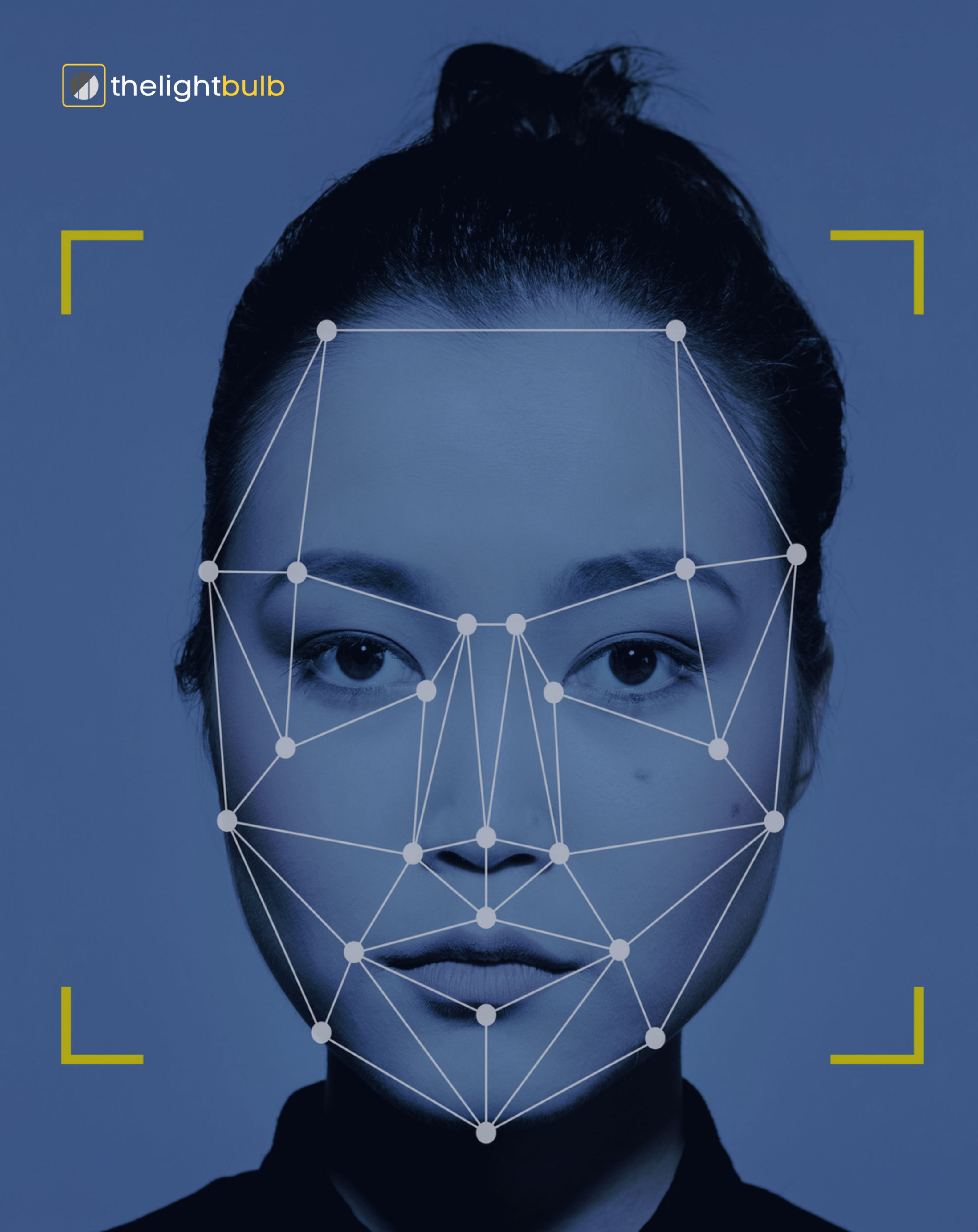
thelightbulb



# Facial Coding

## Accuracy Document

# Table of **Contents**

# Table of **Contents (Contd.)**

# 1. Introduction

Computer vision is a dynamic field that empowers computers to glean valuable insights from digital images and videos, replicating human visual perception through algorithm development. An essential facet of this field is face detection, which involves identifying and localizing human faces within visual data by analyzing distinct features like edges and patterns. Face detection finds diverse applications in security systems, surveillance, biometrics, and photography, contributing significantly to these domains.

Within the realm of computer vision, emotion detection is a compelling subfield that delves into the interpretation of facial expressions to discern underlying emotions. This analysis relies on the Facial Action Coding System (FACS), which assigns specific action units to facial muscle groups, enabling accurate inference of emotions displayed. Emotion detection with FACS boasts wide-ranging applications in market research, psychology, human-computer interaction, and entertainment, providing valuable insights into human emotions and behavior

# 2. Background

## 2.1 Machine Learning

Facial emotion classification using machine learning is a captivating field that employs sophisticated algorithms and techniques to analyze and decipher human emotions from facial expressions. By training models on vast datasets of labeled facial images, machine learning algorithms gain the capability to detect and categorize emotions such as happiness, sadness, anger, fear, and surprise, leveraging distinctive patterns and features present in each expression. This technology finds a multitude of applications, encompassing enhanced human-computer interaction, augmented emotional intelligence in artificial systems, and domains like healthcare, customer service, and entertainment, where understanding emotions plays a crucial role.

The implications of unlocking the capacity to comprehend and respond to human emotions are significant, as machine learning for facial emotion classification paves the way for more empathetic and intuitive interactions between humans and machines. This breakthrough enables technology to be more attuned to human needs, feelings, and preferences, fostering a more harmonious and meaningful integration of machines into our daily lives.

## 2.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) stand out as robust tools for facial emotion classification due to their ability to automatically extract meaningful features from images. They are particularly adept at analyzing facial expressions, making them a perfect fit for this task. When applied to grayscale or color face images, CNNs excel at identifying patterns and structures associated with various emotions, enabling accurate emotion recognition and classification. This advanced technology holds immense potential for improving emotion-related applications in fields such as human-computer interaction, healthcare, and psychology, among others.

Convolutional Neural Networks (CNNs) are multi-layered architectures comprising convolutional, pooling, and fully connected layers. They extract local features through convolutional layers, down-sample feature maps via pooling layers, and perform classification based on high-level features with fully connected layers. CNNs revolutionize computer vision tasks like facial emotion classification by automatically learning relevant features from raw image data, eliminating the need for manual feature engineering and achieving impressive performance in various image-related applications.

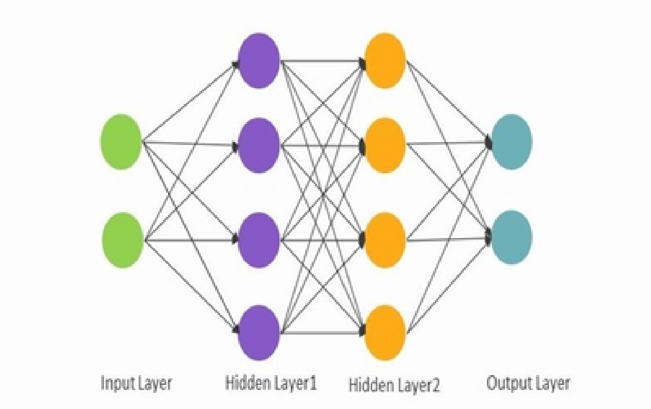


**Fig. 2.1 Simple neural network**

In the training phase, a CNN is fine-tuned using a labeled dataset of facial images, employing techniques like backpropagation and stochastic gradient descent to adjust its parameters. The objective is to minimize the discrepancy between the predicted emotion labels and the actual ground truth labels. Once trained, the CNN becomes capable of classifying facial emotions by analyzing an input image and generating a probability distribution across different emotion categories. Higher probabilities correspond to recognized emotions, leading to accurate classification results. CNNs have demonstrated exceptional performance in facial emotion classification, surpassing previous benchmarks. Their remarkable ability to learn discriminative features from facial images establishes them as indispensable tools for comprehending human emotions through visual cues.

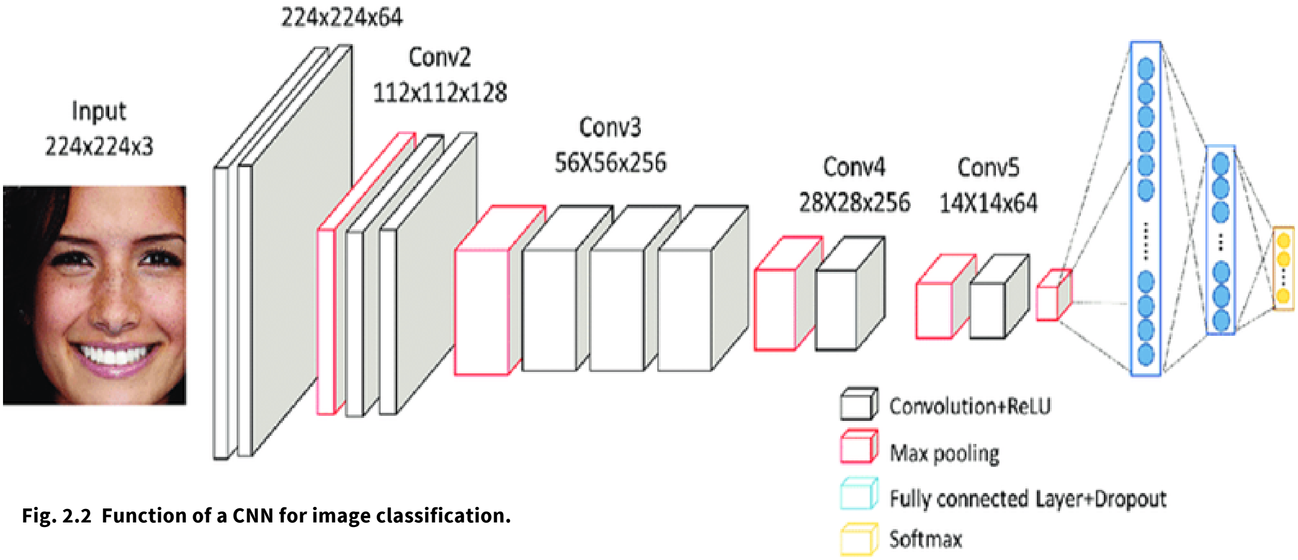


**Fig. 2.2 Function of a CNN for image classification.**

## 2.3 Facial Emotion Analysis

Facial emotion recognition assumes a critical role in customer behavioral research, yielding indispensable insights into emotional responses and preferences. Through meticulous analysis of customers' facial expressions, researchers can effectively shape marketing strategies and elevate the overall customer experience. This technology finds diverse applications in Advertisement Testing, Product Testing, UX Testing, Customer Satisfaction Analysis, and Market Research, where it provides a non-intrusive and objective means of acquiring real-time data. Consequently, businesses can leverage this valuable information to craft personalized experiences and make data-driven decisions in their marketing and customer engagement strategies.

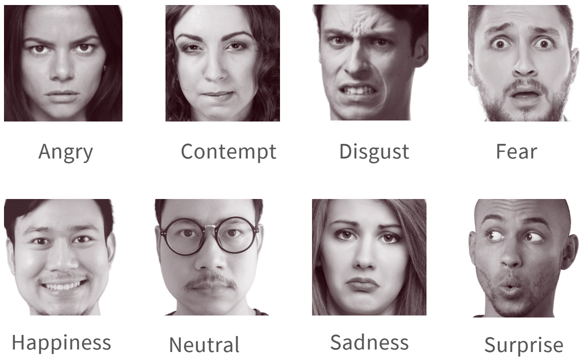


**Fig. 2.3 Faces representing various emotions**

## 2.4 Facial Expression Recognition (FER) Dataset

The FER2013 dataset is a widely used dataset for facial emotion recognition. It contains grayscale face images with emotion labels, widely used for facial emotion recognition. Here is the metadata information for the FER2013 dataset:

**Dataset Name:** FER2013 Source: The dataset was created by Pierre-Luc Carrier and Aaron Courville at the University of Montreal.

**Data format:** The dataset is in CSV format, with each row representing an image and having two columns: The "emotion" column holds the emotion label (0-6), while the "pixels" column stores the image's pixel values as a space-separated string.

## 2.5 Traditional Vs TheLightBulb's approach

Traditional emotional analysis algorithms are limited by CNN models. Our network combines base models (VGGNET, Xception, etc.) with additional CNN and dense layers, optimizing performance and enabling excellent generalization and real-time data processing.

| | Traditional Approach | LB Approach |
|---|---|---|
| Transfer learning | No | Yes |
| Datasets for validation | Limited | Multiple |
| Generalization on new data | Limited | Good |

# 3. Implementation

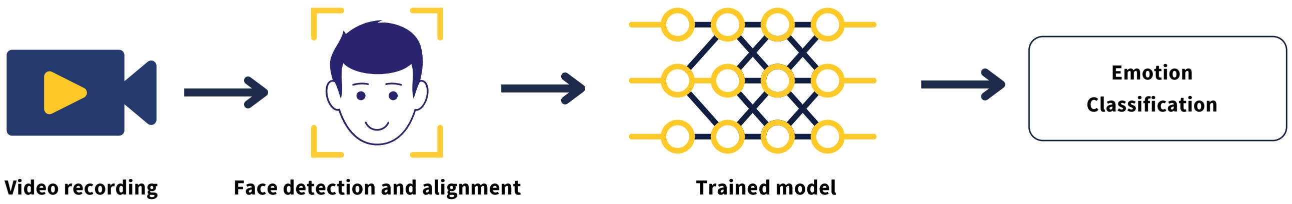


**Fig. 3.1 An overview of the implementation of Facial Encoding product:**
**1. Data acquisition 2. Data-preprocessing 3. Model Prediction 4. Classification.**

## 3.1 Data Pre-processing

Data preprocessing in facial emotion analysis frequently employs the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm, a deep learning-based method designed to detect and align faces for input into an emotion classification model. MTCNN comprises three neural networks, namely P-Net, R-Net, and O-Net, which operate in a cascaded manner to detect faces through diverse image regions and scales, leading to bounding box identification. Following detection, MTCNN performs facial landmark detection to ensure consistent face alignment, normalizing positions, orientations, and scales, effectively mitigating variations caused by head pose, expressions, and occlusions.

Upon detecting and aligning faces, the data undergoes cropping and resizing to attain a fixed size compatible with the emotion classification model. Additionally, grayscale conversion is often applied to reduce computational complexity and eliminate irrelevant color variations, simplifying subsequent processing. Moreover, data augmentation techniques, such as random rotations, translations, and flips, are employed to enhance the diversity of the training data. These variations enhance the model's robustness to diverse poses, lighting conditions, and facial expressions.

In summary, MTCNN-based data preprocessing encompasses face detection, facial landmark detection for alignment, cropping, resizing, grayscale conversion, and potentially data augmentation. These comprehensive steps ensure standardized, aligned, and optimized input data, facilitating accurate emotion classification using subsequent machine learning models.

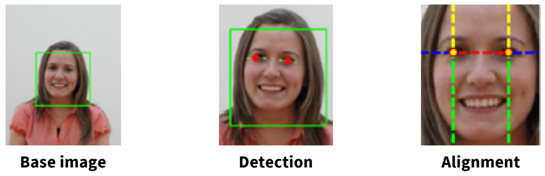


**Fig. 3.2 Faces are detected from images and aligned correctly.**

**Source: CMC-Computers Materials & Continua, vol. 67.

## 3.2 Transfer Learning

Transfer learning represents a potent technique in machine learning and deep learning paradigms, specifically for facial emotion classification. This approach leverages pre-trained models, such as VGGNet, ResNet, or InceptionNet, initially trained on related tasks like ImageNet, to serve as a starting point for training a new model on facial emotion classification. The pre-trained model functions as a feature extractor, where early layers learn low-level features, and later layers capture higher-level features.

To adapt the model for facial emotion classification, the final layers are modified to align with the number of emotion classes. These layers undergo training using a new dataset of labeled facial images dedicated to emotion classification, while the weights of early layers are either kept fixed or fine-tuned with a small learning rate. By exploiting transfer learning, facial emotion classification models effectively capitalize on learned representations from large-scale datasets, even when dealing with limited labeled data. The application of transfer learning leads to notable improvements in performance, convergence speed, and significantly reduces the demand for extensive training. Consequently, transfer learning emerges as a pivotal method to enhance the effectiveness and efficiency of facial emotion classification models.

| Model | LB Approach |
|---|---|
| Xception | **79%** |
| DensNet | **77%** |
| Inception | **75%** |
| VGGNet | **78%** |

**Table 3.1 Accuracy of pre-trained models on Imagenet dataset.**

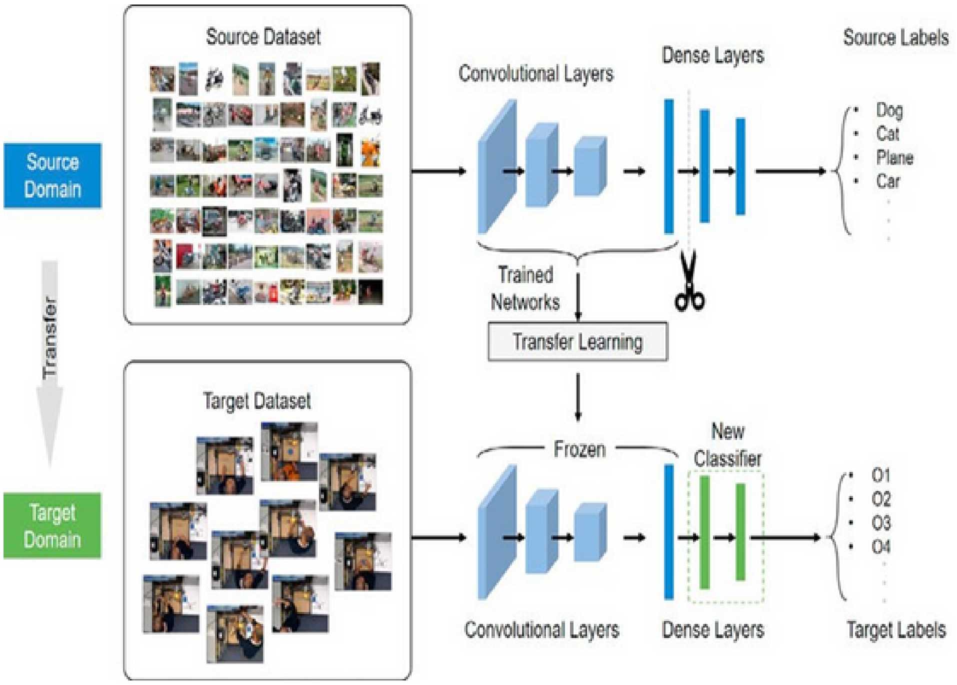


**Fig. 3.3 Transfer learning helps to utilize the knowledge learnt from millions of images (Source Dataset) to train our network with limited number of images(Target Dataset) .**

**Source: researchgate.net.

### 3.3 Network Training

**Model Architecture:** Our emotion classification task employs a modified architecture using a pre-trained CNN model like Xception or VGG16. The base model's weights are loaded from pre-trained models trained on datasets like ImageNet. Additional layers are added for face emotion classification customization.

**Data Generation:** Our code utilizes Keras' ImageDataGenerator for efficient loading, preprocessing, and augmentation of training and validation data.

**Model Training:** The model is trained using an optimizer (Adam or SGD) and a loss function (categorical cross-entropy). The fit() method is used with training data generator, specifying steps per epoch and validation. Evaluation occurs with the validation data generator.

**Model Evaluation and Fine-tuning:** After training, the model's performance is assessed using metrics. The code saves the best model based on validation loss and allows fine-tuning options for further optimization.

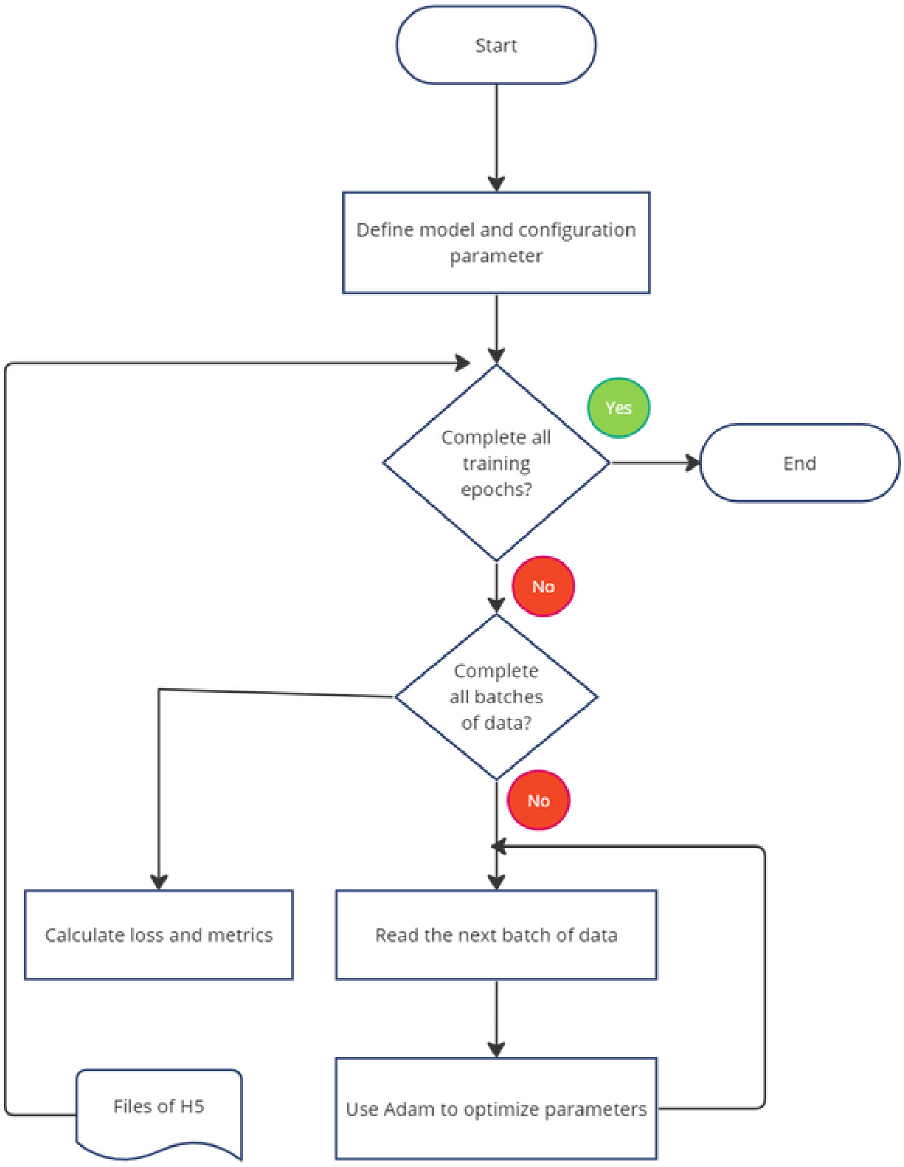

Start
Define model and configuration parameter
Complete all training epochs?
Yes
End
No
Complete all batches of data?
No
Calculate loss and metrics
Read the next batch of data
Files of H5
Use Adam to optimize parameters

**Fig. 3.4 Transfer learning helps to utilize the knowledge learnt from millions of images(Source Dataset) to train our network with limited number of images (Target Dataset).**

**Source: Procedia Manufacturing, vol. 48, pp. 926– 931, 2020

# 4. Implementation

Accuracy compares predicted results of Computer Vision models with emotions tagged by specialists for training and evaluation of pre-tagged images.

### Defining Accuracy

The **percentage match** between the emotions predicted by the technology and the emotions recognized by expression specialists for the same visual stimuli is defined as accuracy. All images are pre-tagged with corresponding emotions by the specialists.

### Stimulus

We compared the results on a statistically significant dataset for each emotions. These were pre-tagged by facial expression experts with an inter-rater reliability (degree of agreement among independent observers) of min 70%.

### Environment

We conducted the study in **real-world settings**, using data of actual non-ideal situations for accuracy testing. The accuracy **results will be higher** in ideal conditions and controlled environments.

## 4.1 Accuracy Results

**LB Model v1** has an **overall accuracy of 76%**, and Thelightbulb's **LB-AWS Hybrid Model** has an **overall accuracy of 92%**, as shown in Table 4.1. Furthermore, our hybrid model has the **highest accuracy for five emotions:** Neutral, Happiness, Contempt, Anger, and Sadness. All emotions are predicted with greater than 85% precision.

| Emotion | LB Model | LB AWS Hybrid Model |
|---|---|---|
| Anger | 77% | 93% |
| Contempt | 67% | 91% |
| Fear | 63% | 81% |
| Happiness | 96% | 96% |
| Neutral | 79% | 95% |
| Sadness | 72% | 95% |
| Surprise | 80% | 94% |
| **Overall** | **76%** | **92%** |

**Table 4.1 Accuracy of LB models on test data**

## 4.2 Claimed Accuracy

Anger, contempt, fear, joy, neutrality, sadness, and surprise were all correctly identified with an **overall accuracy of 76%** for our **LB Model v1** and **92% for LB-AWS Hybrid Model.**

# 5. Facial Coding Insights

## 5.1 Engagement Levels

Engagement levels are metrics that track how actively the audience is involved with the stimuli. The engagement level is used in analyzing the efficacy of the content and how people respond by interacting with videos, workshops, etc. Its rated out of 100. The higher the score the better.

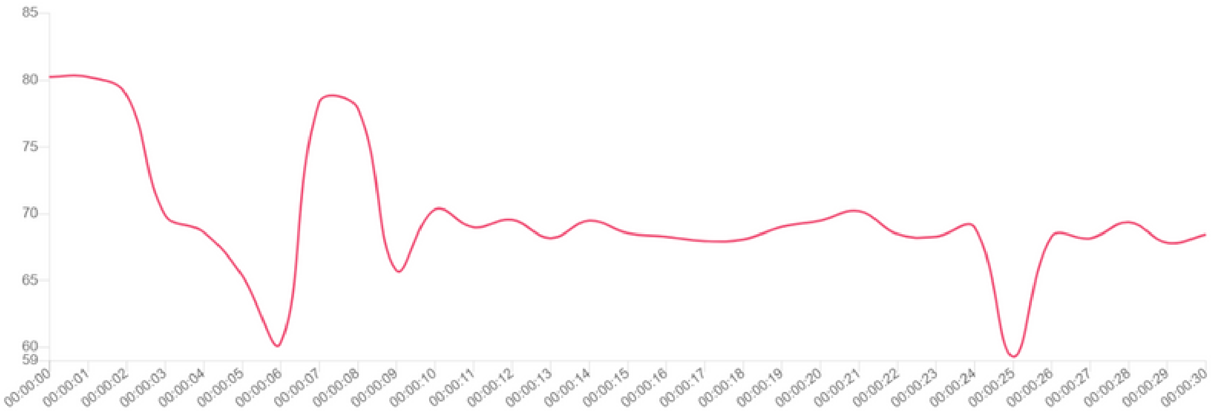


**Fig. 5.1 Average attention and engagement levels of all testers for the given stimulus sec-by-sec.**

## 5.2 Emotion Distribution

Emotion Distribution is metric that tracks the facial muscle movement or action units (AU) that correspond to a displayed emotion in response to the active involvement of the audience with the content. All the emotions together sum upto 100. For positive emotions, the higher the score the better and for negative emotions, the lesser the score the better.

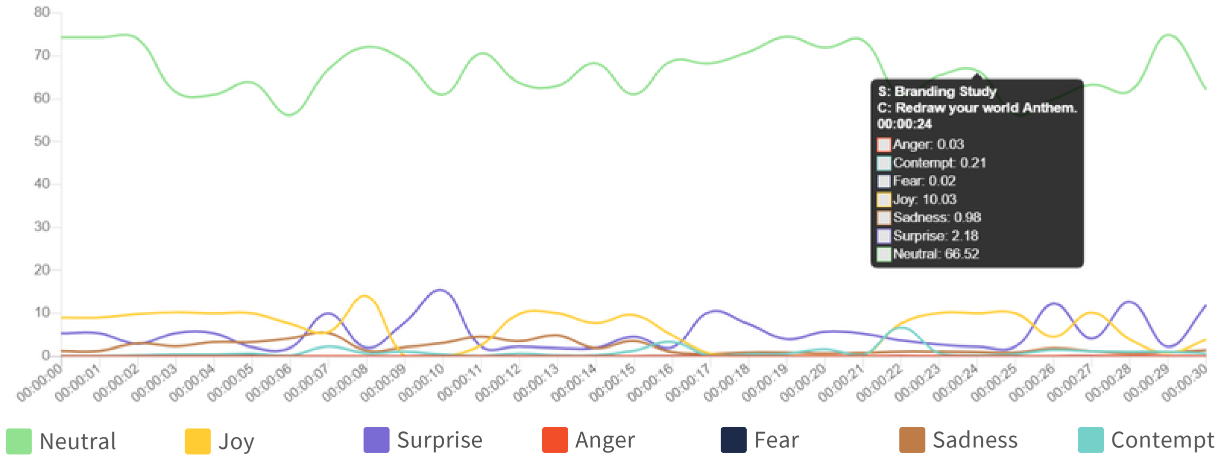
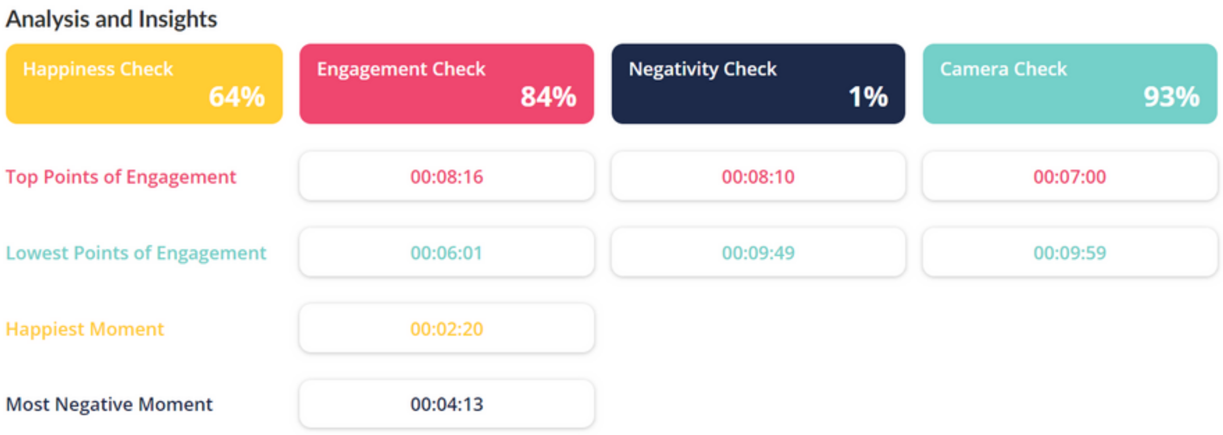


**Fig. 5.2 The average emotional response of all testers for the given visual stimulus on a sec-by-sec basis.**

## 5.3 Analysis and Insights

Analysis and insights highlight the overall emotion check along with peak emotional response for the given visual stimulus marked by time-stamp.



Analysis and Insights

| | | | |
|---|---|---|---|
| Happiness Check<br>64% | Engagement Check<br>84% | Negativity Check<br>1% | Camera Check<br>93% |
| Top Points of Engagement | 00:08:16 | 00:08:10 | 00:07:00 |
| Lowest Points of Engagement | 00:06:01 | 00:09:49 | 00:09:59 |
| Happiest Moment | 00:02:20 | | |
| Most Negative Moment | 00:04:13 | | |

## 6. Understanding **Accuracy and Errors**

### 6.1 How Accurate is Facial Coding

Facial coding systems must be supplemented with other building blocks. It is impossible to provide a universally applicable estimate of accuracy for the system intended to deploy. Companies might share accuracy as measured by public benchmark competitions, but these accuracies are dependent on its details and will not be the same as the accuracy of a deployed system.

Ultimately, system accuracy is determined by a variety of factors, such as the technology and its configuration, environmental conditions, use cases, how people interact with the camera and interpret the output.

| | |
|---|---|
| True positive or true accept<br>👤 ✓ 👤 | The emotion in the probe image is compared against labelled dataset and their emotions are correctly matched. |
| True negative or true reject<br>👤 ✗ 👤 | The emotion in the probe image is compared against labelled dataset and they are not matched. |
| False positive or false accept<br>👤 ✓ 👤 | Either the emotion in the probe image is not labelled in the dataset but is matched to an labelled emotion OR the emotion in the probe image is labelled but is matched with the wrong emotion. |
| False negative or false reject<br>👤 ✗ 👤 | The emotion in the probe image is labelled in the dataset, but they are not matched. |

A facial coding system's accuracy is determined by a combination of two factors:

- How often the system correctly identifies an emotion that is annotated in the system and
- How often the system correctly finds no match for the emotion which is not annotated or labelled.

These two "true" conditions, along with two "false" conditions, combine to describe all possible outcomes of a facial recognition system.

## 7. Best Practices for **Improving Accuracy**

Facial coding technology is improving, and many systems, such as Thelightbulb.ai's ML Model, can perform well even in less-than-ideal conditions. However, there are specific steps you can take to ensure that the facial coding system produces high-quality results.

### Plan for Evaluation Phase

Before deploying or rolling out any facial coding system on a large scale, It is advised to the system owners to conduct an evaluation phase in the context where the system will be used and with the people who will interact with it.

Work should be done with analytics and research teams to collect ground truth evaluation data to:

- Establish baseline accuracy, false positive, and false negative rates.
- Choose an appropriate match threshold to meet the objectives.
- Determine whether the error distribution is skewed toward specific groups of people.

This evaluation should take into account the deployment environment and any variations within it, such as lighting or sensor placement, as well as ground truth evaluation data that reflects the diversity of people who will interact with the system.

To help tune the system and ensure successful engagement, in addition to telemetry data (collection and analysis of real-time facial expressions during the process of facial coding), you may want to analyze feedback from people making decisions based on system output, satisfaction data from people who are subject to analysis, and feedback from existing customer voice channels.

### Meet Image Quality Specifications

Image quality is critical for accurate facial coding, so make certain that both the images used to enroll and the probe images meet the following requirements:

- Full-frontal head and shoulder view without obstruction.
- Face size is at least 200x200 pixels with at least 100 pixels between the eyes. Faces are detectable when their size is as small as 36x36 pixels, but for best performance, We recommend a minimum size of 200x200 pixels.
- Enroll multiple images of each emotion. Include images that represent typical variations in how the person’s face appears to the camera, for instance, with and without glasses, from different angles.

## Control Image Capture Environment

**Lightening and camera calibration:** Examine how well the detail of people's faces can be seen in images taken with the camera planned to use.

- Camera images in proper lighting conditions. Is the lighting too bright or too soft? Are people's faces backlit? Is there too much light from one side and not enough from the other? Place sensors away from areas with harsh lighting whenever possible.
- Is the lighting adequate to capture the details of people's faces with varying skin tones?

**Backgrounds:**

- Strive for backgrounds that are neutral and non-reflective. Avoid backgrounds with faces, such as those with pictures of people or where people other than the person to be recognized are prominent in the photo.

**Sensor placement and maintenance:**

- Position sensors at face level to best capture images that meet the quality specifications.
- Ensure sensors are regularly checked for dust, smudges, and other obstructions.

## Plan for Variations in Subject Appearance and Behavior

**Facial occlusions:** When a person's entire face is visible, facial coding works best. Faces can be partially or completely obscured for a number of reasons, including:

- Religion: Headwear that covers or partially obscures faces.
- Weather: Garments like scarves wrapped across the face.
- Injury: Eye patches or large bandages.
- Vision Disability: Very opaque glasses and pinhole glasses (other glasses and lenses should be fine).
- Personal style: Bangs over eyebrows, baseball caps, large facial tattoos, etc.

**Subject Behavior:** When subjects are not facing the camera, occluding their face with their hands (such as brushing hair out of their eyes), moving too quickly for the sensor to capture their image, or their expression is extreme, image quality may suffer (like yawning widely with their eyes closed). To address these issues:

- Design the user experience so people understand how to provide high-quality images.
- Create an environment where people naturally face the camera and slow down.
- Provide clear instructions for how people should behave during analysis (eyes open, mouth closed, sit still, etc.).

## Design the system to support human judgment

Meaningful human review is important to:

- Detect and resolve cases of misidentification of emotions or other failures.
- Provide support to people who believe their results were incorrect.
- Identify and resolve changes in accuracy due to changing conditions (like lighting or sensor cleanliness).

The user experience created to support the people who will use the system output should be designed and evaluated with those people to understand how well they can interpret the output, what additional information they might need, how they can get answers to their questions, and ultimately, how well the system supports their abilities to make more accurate decisions.

## 8. Frequently Asked Questions (FAQs)

**1.What is the engagement score?**

Engagement score is also called as attention score. This is a function of all emotions demonstrated, head positions. Any frame analyzed is given an Engagement score of 100. Also a separate score comprising of various emotions with various levels of probabilities, the dominant emotion and non-dominant emotions sum up to 100.

**2. What does the engagement score mean? Is it good or bad engagement?**

A higher score represents good customer engagement. Engagement is proxy to people paying attention to stimuli. Any score above 80% is considered good engagement but depending upon context lower or higher scores may be considered good.

**3. How does the tool analyze facial expressions and eye movements? Can you provide an overview of the underlying technology?**

The tool uses AI technology to analyze facial expressions.

**4. What emotions does the tool detect and track? Does it cover a wide range of emotions or specific ones?**

Joy, Fear, Sad, Anger, Contempt, Surprise, Confusion and Neutral.

**5. Can you explain how the tool measures head position and why it is included in the calculation? How does head position relate to engagement?**

The tool can extract features which can detect if the faces are rotated in the videos using MTCNN model. The head position is correlated with customers attention.

**6. What sample size do you suggest for facial coding and eye tracking?**

The tool can handle any file size typically coming from an normal video camera.

**7. In which type of study do we get individual results and aggregate results?**

In Ad testing, concept testing, and content testing, we can obtain both aggregate and individual results when considering that all users are exposed to the same stimuli. However, in the case of website UI/UX testing, we only receive individual results because each user has a unique browsing journey, which poses a challenge in gathering aggregate results.

**8. How does your tool segregate the diversity in India? For example, North vs. South.**

It is possible upon request to detect the race using Deep Face based on CNN models.

**9. How does the tool analyze facial expressions? Can you provide an overview of the underlying technology?**

Facial emotion analysis technology utilizes computer vision and machine learning algorithms to detect and interpret emotional expressions from facial images or videos.

**10. What emotions does the tool detect and track? Does it cover a wide range of emotions or specific ones?**

The tool analyzes the facial emotions of the person (Joy, Fear, Sadness, Anger, Contempt, Surprise and Neutral).

**11. Can the tool differentiate between genuine and fake expressions? How accurate is it in detecting emotions?**

This task is in progress and we are currently creating such datasets.

**12. How does the tool handle different lighting conditions or variations in video quality? Does it require specific camera setups?**

The tool can handle low resolution videos under reasonably less lighting conditions.

**13. What kind of data outputs does the tool provide? Are there detailed reports or visualizations available?**

The tool can provide the probabilities for the emotions and we can find the emotion with highest probability score.

**14. What does ”real-time emotions” mean? Can the researcher have immediate access to individuals’ facial coding?**

On completion of the stimuli by one respondent - 15 min (95% of the time) but aggregation of results may take time depending upon number of respondents and the complexity of the study.

**15. How much time does the tool take to process the data and send the output?**

The tool takes approximately 15 min to process the data depending upon number of respondents and the complexity of the study as mentioned above.

## 9. Appendix

### 9.1 Note On Comparison Strategy

- Instead of choosing standard validation sets, we wanted to test the algorithms in real-world scenarios with data from actual respondents. Most of the generic algorithms from other systems are trained on curated data of facial emotions which are either posed or selected carefully by picking the high-intensity emotion. Almost all the systems are comparable on such data sets and give more than 90% accuracy.
- Our model was trained on in-real-world images which are pulled from real video-watching sessions with relaxed conditions on light and pose. The algorithm uses deep learning-based algorithms to learn from such datasets.
- We validated the new model by giving it random images from our validation set which is curated from video-watching sessions.
- Since not all models give the same emotions and some have failure issues for some images, we recommend looking at accuracy numbers for each emotion for a better comparison than looking at the overall number.

Thelightbulb.ai uses latest in computer vision, emotion ai & machine learning technologies to measure attention and emotions of opt-in participants as they consume content & experiences online.

**LET'S TALK!**

SALES@THELIGHTBULB.AI