**Task 3: - Database: Table 1(id, name, embedding) Table 2(id, checking, date)**

**- CRUD operations**

**CReate, remove, update, delete**

**- tạo dữ liệu cho database( lấy ảnh của thành viên nhóm....)**

**- documents(làm những feature nào, .....)**

**Dataset: CASIA-WebFace**

**Link:** [**CASIA-WebFace**](https://www.kaggle.com/datasets/debarghamitraroy/casia-webface/data)

## Overview:

The paper focus on improving face recognition performance through deep convolutional neutural networks (CNNs) trained on very large-scale datasets. CASIA-WebFace contains approximately 10,575 subjects and 494,414 images.

## Creation of CASIA-WebFace Dataset:

* A large-scale face dataset built using a semi-automated process that collects and annotates images from IMDb.
* Ensures no overlap with the Labeled Faces in the Wild (LFW) dataset for unbiased evaluation.
* The dataset provides a new standard for training and evaluating face recognition algorithms in the wild.

## Development of a Baseline Deep CNN:

* The authors trained a deep CNN with **11 layers**, integrating popular techniques like ReLU activation, dropout, and a combination of identification and verification loss functions.
* The network's architecture includes 10 convolutional layers, 5 pooling layers, and 1 fully connected layer.

## Evaluation on Benchmark Datasets:

* Performance on **LFW** and **YouTube Faces (YTF)** datasets shows that the proposed approach achieves state-of-the-art results.
* On LFW:
  + The single-network model achieves **97.73% accuracy**, outperforming Facebook's DeepFace and comparable to DeepID2 ensembles.
* On YTF:
  + The model achieves **92.24% accuracy**, slightly surpassing DeepFace.

## Methodology:

**Dataset Collection:** Images were crawled from IMDb, leveraging structured metadata for efficient annotation. The authors employed a clustering method to annotate identities and manually validated the results.

**Training Pipeline:**

* Images were aligned and normalized to 100x100 grayscale inputs using facial landmarks.
* A small-filter, deep architecture was employed to reduce parameters and improve generalization.
* Augmentation techniques like mirroring were used to increase robustness.

# DATABASE

A screenshot of a computer

Description automatically generated

## 1. Geometric Features

These focus on the spatial relationships and key points on the face. Common features include:

* Eye position and shape
* Eyebrow position and curvature
* Nose length, width, and shape
* Mouth position, size, and shape
* Distance between key landmarks (e.g., inter-eye distance, nose-to-mouth distance)
* Facial contour (jawline and face shape)

Typically, modern systems use around **68 to 128 key landmarks** as a baseline for geometric analysis.

## 2. Appearance-Based Features

These analyze the texture and patterns of the face. Techniques include:

* Skin texture (wrinkles, pores)
* Color patterns (e.g., tone differences in skin)
* Surface curvature (3D face scans)

Appearance-based systems rely on pixel-level data or learned features from images.

## 3. Feature Extraction Techniques

Modern systems extract a mix of features using machine learning, particularly deep learning:

* **Principal Component Analysis (PCA)**: Used in Eigenfaces to reduce the dimensionality of face data.
* **Linear Discriminant Analysis (LDA)**: Enhances separability between individuals.
* **Convolutional Neural Networks (CNNs)**: Extract features like edges, contours, and regions of interest hierarchically.