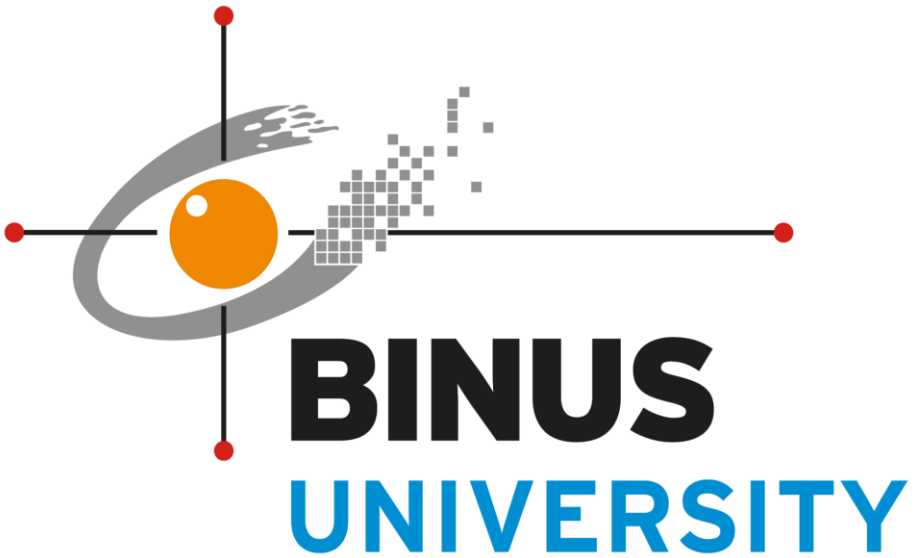
Vision-Based Classroom Attendance System using Feature Extraction and SVM Classifier



Brian Alexander - 2702282351
Jason Christian Budhihartono - 2702326593
Melvern Michio Chie - 2702220946

Kelas : LE01

Mata Kuliah : Computer Vision

Dosen Pengampu:

Dr. NUR AFNY CATUR ANDRYANI, S.Si., M.Sc.

Tahun Ajaran 2025/2026

1. ABSTRACT

Our group built a face recognition system that pairs deep-learning feature extraction with a standard linear Support Vector Machine (SVM) to see if we could maintain high accuracy with a lighter classification layer. Our tests showed a notable performance gap between the models: InsightFace hit a perfect 100% accuracy mark, whereas Face.evoLVe trailed behind at 92% on our dataset. These results suggest that even when using a "heavy" feature extractor, a simple SVM classifier is more than capable of handling the final identification task efficiently. This hybrid setup proves you don't necessarily need an end-to-end deep neural network to get professional-grade results.

2. INTRODUCTION

Advancements of artificial intelligence in the past decade have transformed how machines interpret the world. Modern deep learning models, especially the ones in computer vision, can now reliably detect, classify and describe visual content with accuracy that once seemed unattainable. These capabilities have enabled more practical applications in various fields such as medical and industrial fields

Within education, administrative tasks such as attendance taking remain time-consuming and are prone to errors when done manually. Traditional roll calls are prone to errors and time consuming, meanwhile Wi-Fi or phone check-in systems can be bypassed by students who are not physically present, and card-based methods are easily exploited when cards are tapped by someone else. These limitations show the need for a securely verifiable attendance method.

In this project we introduce an automated attendance system that uses face recognition technology to identify students and record their presence. The system performs face detection, alignment, and feature extraction before then having the images be processed by a SVM model which has been previously trained on the faces.

3. RELATED WORK

A two stage pipeline is frequently used in recent works: deep face models are used to extract high-dimensional face embeddings that represent their face feature andE structure, then those embeddings will be used to train classical machine learning classifiers for final classification and prediction. This pipeline is efficient and light-weight because it isolates the computational heavy feature extraction from relatively light classification stages suitable for realtime systems, it also supports small datasets as machine learning classifiers need less dataset compared to other deep learning models.

FaceNet, Insightface, and face.evoLVe are several deep face models that have been widely used because of their powerful discriminative skills that are able to produce accurate face features. Several studies demonstrate that combining SVM with those deep face models is able to achieve highly competitive results and is able to integrate with realtime systems. Particular research is implementing FaceNet deep face model with SVM for realtime systems that have reported to achieve an excellent performance with 98% accuracy in evaluation and being able to differentiate faces correctly in the systems [1]. Those findings highlight SVM’s capability in capturing well-structured embedding decision boundaries between each class.

In addition to identity recognition, a study evaluating Insightface embeddings on a gender classification task and comparing several machine learning classifiers. This study finds that SVM achieved the highest accuracy (96%) among the other models and demonstrates that SVM is a solid choice for classification tasks that involve high quality features [2]. Those findings encourage us to explore those deep face models with SVM to produce a quick, light-weight, and accurate model that is suitable for face-based attendance systems.

4. METHODOLOGY

4.1 Dataset

The dataset used in this particular research is all of the researcher's face, the amount of each face is limited to less than five to test how powerful each face feature extractor model is able to produce an accurate face embeddings for SVM classifier. Data augmentation is applied to produce variation of images between each class to further enhance the model capability at capturing face structure variation from each class.

4.2 Feature Extraction

This research used the combination of feature extraction with traditional machine learning models to produce discriminative facial feature embeddings with pre-trained face recognition models and are later used as input for machine learning models. All of the deep face recognition models produced 512-D of vector embeddings capturing the detailed facial structure and features for each face image.

4.2 Traditional vs Modern Feature Extraction

The main differences of traditional face feature extraction and modern face is on how they produce the embeddings. Traditional feature extractor rely on simpler features such as edge, corner, gradient, and geometry which doesn't learn identity but predefined pattern. Meanwhile modern feature extractor learns features by hierarchical means, meaning they learn simpler features such as edge, texture at the early layer, facial part (eye shape, distance) at the middle layer, and at the deep layer they learn unique and specific identity from faces, all of the layer will then be combined into a complete vector embeddings representing person's detailed facial feature. Modern feature extractor not just learns how the person looks, but it learns about who belongs to this face. Modern facial feature extractor also uses the loss function to help the model differentiate each person's facial features by minimizing the distance between the same person's face and maximizing the distance of the different person's face making this extractor highly discriminate towards different face features.

4.4 FaceNet

FaceNet is trained with a deep convolutional neural network (CNN) to optimize the embeddings, it will map the face image to a compact euclidean space where the distances correspond to the face similarity. FaceNet differs from traditional "hand-engineered" features from old Computer Vision techniques as it utilizes convolutional technique and learns features from raw pixels of the face and followed by L2 normalization to produce complex embeddings representing each image facial feature. FaceNet also used triplet loss (Achor, Positive, Negative) to maximize DCNN models by ensuring the distance between the Anchor (reference

face) and Negative (different person's face) is larger than the Anchor and Positive (Different image of the same person as Anchor) by a determined margin [3].

4.5 InsightFace

InsightFace used Additive Angular Margin Loss to improve the traditional softmax loss function to learn from highly discriminative facial features. The goal of this loss is to pull closer for the same person samples and push apart for the different person samples on the hypersphere. A huge difference between triplet loss and ArcFace is that triplet loss conducts a sample to sample comparisons while ArcFace conducts a sample to class to ensure the sample is compared to all of the classes rather than single samples. InsightFace used a ResNet backbone without the bottleneck structure to extract the feature and map it into an embedding vector representing each facial feature of the image [4].

4.6 Face.evoLVe

Face.evoLVe is a library and training framework used for face recognition tasks. The system used Multi-task Cascaded Convolutional Networks to detect faces and cropped it, then the images were augmented and sampled to ensure data balance. Face.evoLVe implements various backbones for feature extraction such as ResNet, Improved ResNet, etc. Face.evoLVe also supports various loss functions such as softmax, arcface, and triplet loss [5].

4.7 Support Vector Machine

Support Vector Machine (SVM) is one of the popular machine learning classifiers with the goal of finding an optimal hyperplane that separates data classes in a high-dimensional feature space. While SVM itself is a linear model, it can also handle non linear classification problems using kernel trick by projecting data into higher dimensional spaces. In this particular research, we implemented a linear kernel because of the limitation of the dataset [6].

4.8 Evaluation

To evaluate the model's performance from each feature extraction, we used several evaluation metrics as the measurement on how well each embedding helps the SVM classifier to classify person faces correctly. Accuracy calculate how many prediction is correctly from all the predictions, classification report generates precision, recall, and f1 score for each classes, confusion matrix to visualize the model's performance across all classes, ROC curve to see how confident the model on predicting the correct class, and finally test prediction with new image to see if the model is actually able to recognize person's face outside of the dataset.

5. IMPLEMENTATION & RESULTS

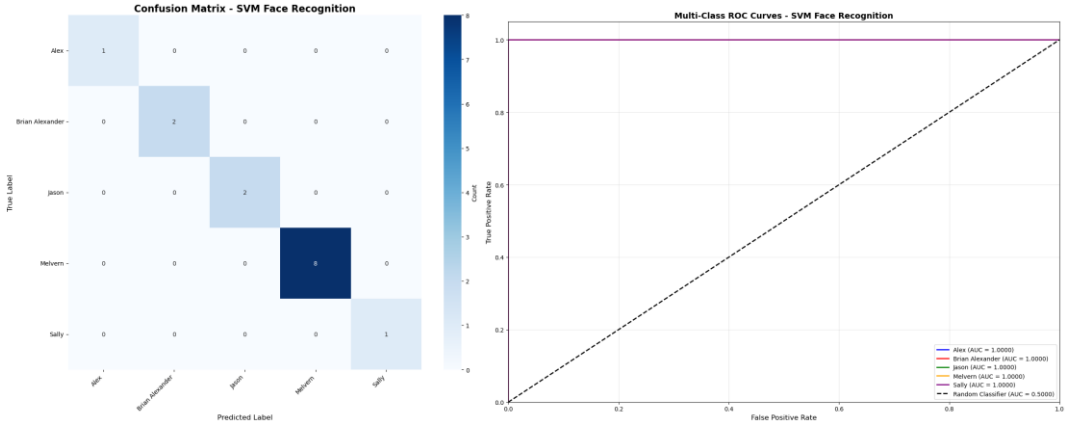
5.1 FaceNet Result

SVM Model Accuracy: 1.0000

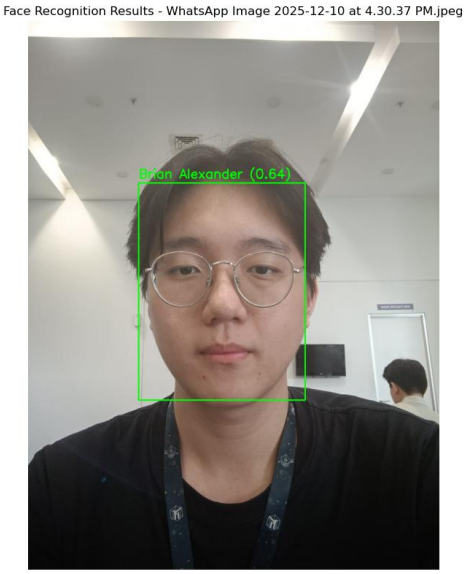
Classification Report:

	precision	recall	f1-score	support
Alex	1.00	1.00	1.00	1
Brian Alexander	1.00	1.00	1.00	2
Jason	1.00	1.00	1.00	2
Melvern	1.00	1.00	1.00	8
Sally	1.00	1.00	1.00	1
accuracy			1.00	14
macro avg	1.00	1.00	1.00	14
weighted avg	1.00	1.00	1.00	14

5.1 Classification Report FaceNet+ SVM classifier



5.1.b Confusion Matrix of FaceNet + SVM classifier; 5.1.c ROC curve of FaceNet + SVM classifier

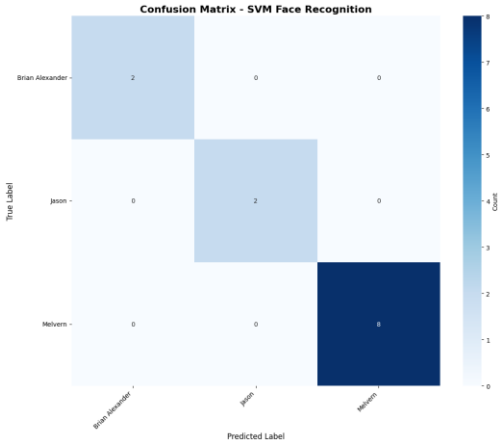


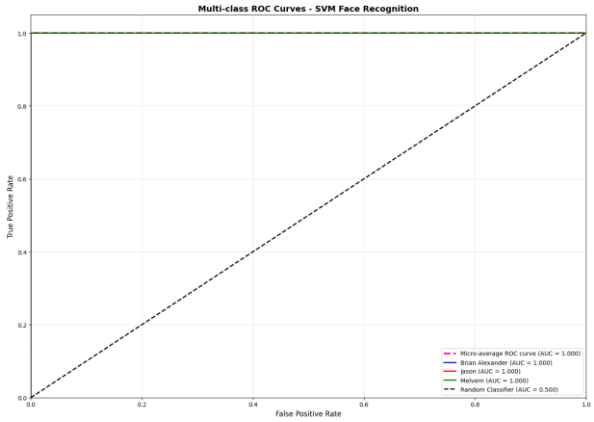
5.1.d Evaluation Test for FaceNet + SVM classifier

5.2 InsightFace Result

```
================================================================================
CLASSIFICATION REPORT
================================================================================

Detailed Classification Report:
                precision    recall  f1-score   support

Brian Alexander     1.0000    1.0000    1.0000         2
         Jason      1.0000    1.0000    1.0000         2
        Melvern     1.0000    1.0000    1.0000         8

      accuracy                         1.0000        12
     macro avg      1.0000    1.0000    1.0000        12
  weighted avg      1.0000    1.0000    1.0000        12
```

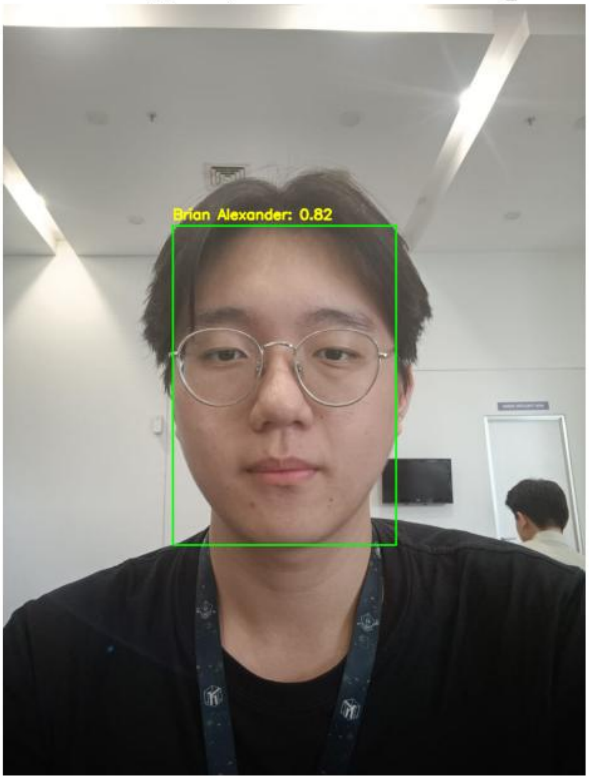
5.2 Classification Report of InsightFace + SVM classifier





5.2.b Confusion Matrix of InsightFace + SVM classifier; 5.2.c ROC curve of InsightFace + SVM classifier

Prediction for WhatsApp Image 2025-12-10 at 16.28.02_cb4337b5.jpg

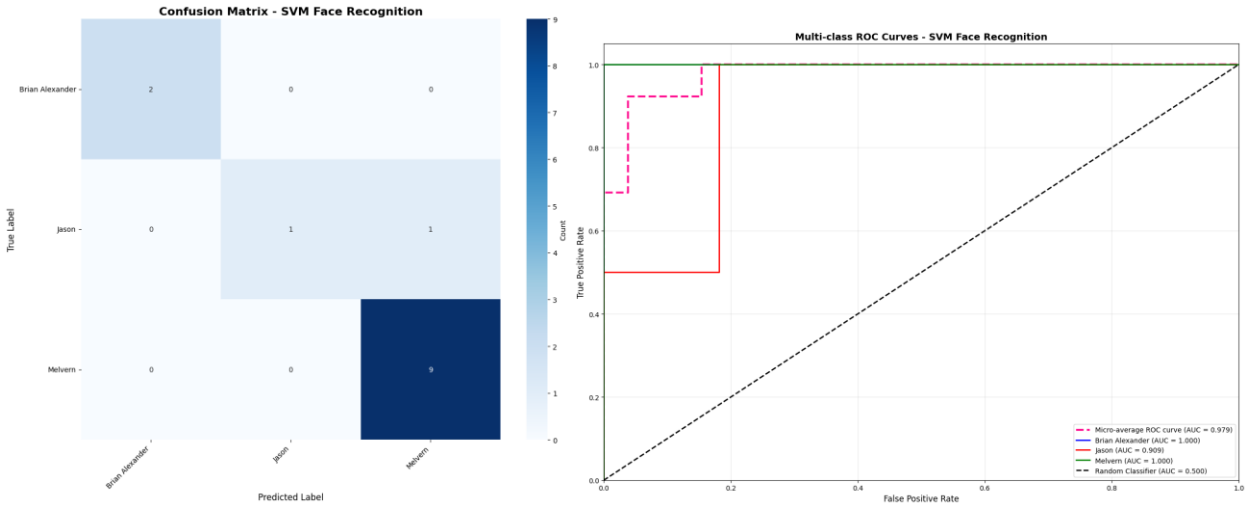


5.2.d Evaluation Test for InsightFace + SVM classifier

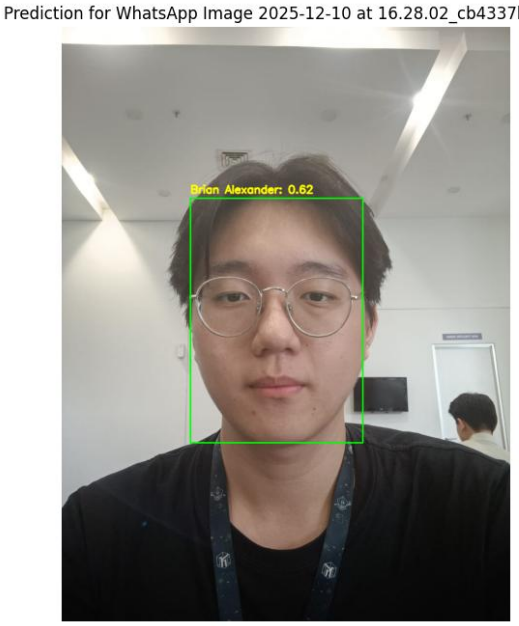
5.3 Face.evoLVe Result

```
================================================================================
CLASSIFICATION REPORT
================================================================================

Detailed Classification Report:
                precision    recall  f1-score   support

Brian Alexander     1.0000    1.0000    1.0000         2
          Jason     1.0000    0.5000    0.6667         2
         Melvern     0.9000    1.0000    0.9474         9

       accuracy                         0.9231        13
      macro avg     0.9667    0.8333    0.8713        13
   weighted avg     0.9308    0.9231    0.9123        13
```

5.3.a Classification Report of Face.evoLVe + SVM classifier



5.3.b Confusion Matrix of Face.evoLVe + SVM classifier; 5.4.c ROC curve of Face.evoLVe + SVM classifier



5.3.d Evaluation Test for Face.evoLVe + SVM classifier

5.4 Implementation

Model with the highest accuracy will then be integrated into a website based attendance system. React is used as the front-end with vite and tailwindcss. The service for student attendance is used express.js with prisma ORM for database modelling. Finally the model itself will be hosted to fastapi for face recognition inference.

6. DISCUSSION & LIMITATIONS

6.1 Discussion

The implementation of FaceNet for facial embedding extraction yielded exceptionally high performance, achieving 100% accuracy on the test set with consistently high confidence scores across all classified images. This indicates that the model successfully learned to distinguish the unique features of the subjects within the provided data, creating distinct clusters in the embedding space. The high confidence levels suggest that the margin between the positive and negative pairs was sufficient for the classifier to make definitive predictions without ambiguity. However, such perfect results should be interpreted with caution, as they likely reflect the constrained nature of the testing environment rather than the model's absolute robustness in broader real-world scenarios.

The primary limitation of this study lies in the small size and limited diversity of the dataset. With a restricted number of samples, the model may have overfitted to the specific lighting conditions, poses, or background features present in this specific collection of images, rather than generalizing to unseen variations. Consequently, while the system performs flawlessly in this controlled context, its performance would likely degrade when introduced to a larger, more heterogeneous population or less ideal capture conditions (e.g., occlusions or extreme angles). Future work should focus on stress-testing the system against a more extensive dataset to better evaluate its scalability and resistance to false positives.

6.1.b InsightFace

The result of the insightface embedding and SVM classifier produced a high accuracy of 100%. This is due to the low number of dataset used as validation, but it also indicates that embedding extracted by insightface is discriminative so the SVM classifier is able to predict faces correctly.

In picture 5.2.b, the model is able to correctly predict all the classes indicating that the model can actually differentiate all the classes' facial features well with insightface feature extraction. Picture 5.2.c also shows that the model not only predicted correctly, but it also confidently predicted the correct result, so the model is not randomly guessing the predictions. Finally evaluation is to test out how well the model’s performance outside of the trained dataset. In the picture 5.2.d, we can see that the model correctly predicted the researcher’s face with a high confidence of 0.81. This truly shows that the model generalizes facial features well across trained classes of researchers.

6.1.c Face.evoLVe

The result of the face.evoLVe embedding and SVM classifier produced a high accuracy of 92%. This indicates that the model can generalize well but slightly worse than the other feature extraction models such as FaceNet and InsightFace even though the dataset is small.

In picture 5.3.b, the model is able to correctly predict almost all the classes indicating that the model can actually differentiate the classes' facial features well with face.evoLVe feature extraction, but still making some mistakes on Jason class indicates the model is struggling to recognize Jason’s facial feature with face.evoLVe embedding . Picture 5.3.c also shows that the model not only predicted correctly, but it also confidently predicted the correct result, so the model is not randomly guessing the predictions. Finally evaluation is to test out how well the model’s performance outside of the trained dataset. In the picture 5.3.d, we can see that the

model correctly predicted the researcher's face with a medium confidence of 0.62 slightly worse than the other two feature extraction. This shows that face.evoLVe as a feature extraction model performs worse than FaceNet and InsightFace.

6.2 Limitation

While the research provides several insights for deep face feature extractor model comparison, there are limitations that need to be acknowledged for future work. The current dataset used for this particular research is too small and diversity between each class is low causing this research not fully capturing the intra-class variance so the overall model performance shouldn't be fully considered as the representative for model's capability. From the side of the algorithm used, this research only implements SVM classifiers, especially linear kernel tricks, which means this research only explores and assumes the feature used is suitable for linear feature space and there is other non linear kernel tricks that can be explored such as RBF and Polynomial. In addition, there is more alternative machine learning architecture or deep learning architecture for classification with the extracted embeddings that might perform better than the current setup used.

7. CONCLUSION & FUTURE WORK

In conclusion, this study successfully demonstrated the feasibility of a hybrid face recognition pipeline for classroom attendance, combining deep learning feature extraction with a lightweight Support Vector Machine (SVM) classifier. The experimental results highlighted a clear performance hierarchy among the tested models, with FaceNet and InsightFace achieving perfect accuracy (100%) on the test set, significantly outperforming Face.evoLVe, which achieved 92% accuracy. These findings suggest that while all three models produce discriminative embeddings, FaceNet and InsightFace provide more distinct decision boundaries for the linear SVM to separate classes effectively. Ultimately, this project proves that leveraging pre-trained "heavy" feature extractors with a simple classical classifier is a highly effective strategy, offering professional-grade recognition capabilities without the need for training complex end-to-end deep neural networks from scratch

To evolve this system from a successful prototype into a robust deployment-ready solution, future work must address the current constraints regarding data and model flexibility. The primary immediate goal is to expand the dataset significantly, increasing both the number of subjects and the diversity of intra-class variations (e.g., different lighting, angles, and occlusions) to rigorously stress-test the model's generalization capabilities. Additionally, while the linear kernel proved sufficient for this limited dataset, further research should explore non-linear SVM kernels, such as Radial Basis Function (RBF) or Polynomial kernels, to handle more complex feature spaces. Finally, evaluating alternative machine learning architectures beyond SVM could uncover even more efficient classification layers, further optimizing the system for real-time processing in busy classroom environments.

8. REFERENCES

[1] L. Q. Vu, P. T. Trieu, and H. Nguyen, "Implementation of FaceNet and support vector machine in a real-time web-based timekeeping application," vol. 11, no. 1, pp. 388–396, 2022, doi: 10.11591/ijai.v11.i1.pp388-396.

[2] M. Farzaneh, "ArcFace Knows the Gender , Too !," pp. 1–9.

[3] F. Schroff and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering".

[4] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, “ArcFace : Additive Angular Margin Loss for Deep Face Recognition,” vol. 14, no. 8, pp. 1–17, 2015.
[5] Q. Wang, *Face . evoLVe : A High-Performance Face Recognition Library*, vol. 1, no. 1. Association for Computing Machinery.
[6] N. H. Ovirianti, M. Zarlis, and H. Mawengkang, “Support Vector Machine Using A Classification Algorithm,” vol. 6, no. 3, pp. 2103–2107, 2022.