Multimodal Approach to Face and Iris Recognition Using Neural Network

1st Zheng Wei, Ng
Computer Science Department
California State University, Fresno
Fresno, USA
loadedwei@mail.fresnostate.edu

2nd Saishnu, Ramesh Kumar
Computer Science Department
California State University, Fresno
Fresno, USA
saishnu_rk@mail.fresnostate.edu

3rd Ren Hao, Wong
Computer Science Department
California State University, Fresno
Fresno, USA
renhao@mail.fresnostate.edu

CSCI 158 - Applied Biometric Security
Professor Amith Kamath Belman
California State University, Fresno
May 10th, 2024

***Abstract*—Biometric recognition systems play an important role in a vast amount of domains that range from security-based applications to identity-based applications. Among the diverse amount of biometric modalities, face, and iris specifically have gained more popularity and interest among the community due to their distinctiveness, universality, and non-intrusiveness. With this, traditional unimodal methods tend to face harsh challenges due to their susceptibility to spoofing attacks and variations within the environment. To combat this, this paper will present an in-depth investigation into the efficacy of a multimodal approach towards face and iris recognition through the means of a neural network. The primary objective of this study is to assess the superiority of multimodal biometric systems in comparison to the traditional unimodal method in terms of security, robustness, and accuracy. By integrating both face and iris data, we aim to manipulate the strengths of these modalities, thereby enhancing the reliability and resilience of a biometric recognition system. Through evaluation, experimentation, and analysis, we can test the performance of the multimodal system under various variations. This research will delve into the neural network architecture and fusion strategy to smoothly integrate face and iris as the two main modalities. Our proposed approach not only facilitates the fusion of biometric information but also ensures the scalability and efficiency to make it suitable for possible real-world deployment. The findings recorded from this study can contribute to the advancement of biometric recognition by providing evidence regarding the superiority of multimodal approaches. In addition, the insights gained from this research can inform the design and implementation of a more robust and secure biometric system, thereby addressing critical challenges in authentication and identification tasks across diverse applications.**

***Index Terms*—Biometric, face recognition, iris recognition, neural network architecture, nearest neighbor, multimodal, unimodal, robustness, accuracy.**

I. Introduction

With the ever-interconnected world we live in today, the need for a robust and secure biometric recognition system has not been more crucial. As our interactions in the digital world increase, there is a need for a reliable authentication mechanism to safeguard sensitive information and provide a strong infrastructure behind the scenes. Biometric authentication can offer a promising avenue to enhance security in various scenarios from access controls to financial transactions. Despite this, the current reliance on the traditional unimodal approach, which solely relies on a single biometric trait, such as a fingerprint, has been demonstrated to be more susceptible to spoofing attacks and environmental variations. Hence, this causes challenges concerning the efficacy and reliability of the authentication mode. With the increase in cyber threats, we would need to shift our minds to develop a more resilient and adaptive form of security.

The concept and idea of introducing a multimodal biometric recognition system is the most suitable solution. By combining multiple biometric modalities, such as face, iris, and fingerprint recognition, multimodal systems can offer an intricate approach to identity verification which enhances both security and accuracy. This shift recognizes the complexity and uniqueness of utilizing human traits therefore reducing the limitations that are associated with the unimodal system thus increasing the robustness of such an authentication approach.

To create and develop such an approach, we would utilize existing technologies today, particularly neural network techniques. Neural networks can imitate the human brain and complex patterns within a dataset and they serve as a cornerstone to developing a more sophisticated biometric recognition model. By using neural networks, researchers can explore different architectures and fusion strategies relating to multimodal approaches.

In this project, we will jump into the efficacy of multimodal biometric recognition with a more specific approach and focus on face and iris recognition. We intend to design and implement a robust form of multimodal recognition system capable of integrating the sourced biometric data. Our objective includes a comprehensive examination of existing multimodal techniques and their overall principles, architectures, and performance metrics. Furthermore, through various experimentations and analyses, we aim to conduct a comparative approach to the advantages of a multimodal biometric system as opposed to an unimodal approach. We also intend to explore the addition of modality weights and fusion to achieve an optimized system that can perform well in enhancing its overall security.

By analyzing its overall efficacy, we hope this research has the ambition to contribute to the advancement of biometric authentication technologies. Moreover, we believe this study can be observed and utilized in real-world applications in a diverse amount of fields. We also intend for the widespread adoption of such an approach as it provides users and consumers more peace of mind.

II. Core Concepts

The core concept of this project is to show how limited an unimodal approach is concerning its vulnerabilities and how implementing a multimodal approach can help strengthen this. We solely focus our attention on utilizing two forms of biometrics that include face and iris through means of implementing a neural network to train and gather the respective data we require. The following are core concepts for this project as they highlight specific sections deemed important in this case:

1. Face Recognition: This process is to identify and verify individuals based on their facial features. It involves capturing facial images and extracting unique features from them such as the distance between their eyes, nose shapes, and contours on their face such as rectangular, diamond, and oval.

2. Iris Recognition: This method uses unique patterns found within an individual's eye to identify themselves. Within the iris itself, it contains various complex patterns of crypt and collarette that are distinctive to each individual and remain constant all the time.

3. Neural Networks: It is a type of machine learning algorithm that contains interconnected nodes that are organized into multiple different layers. Each of the nodes will apply a mathematical operation for its input to pass the results to the next layer. It is also capable of learning complex patterns and performing complex solutions from various datasets hence its importance regarding its use within multimodal biometric systems. Utilizing specific deep learning architectures such as the convolution neural network (CNN) can help with feature extraction, fusion, and classification of multimodal data. Neural Networks can conduct end-to-end learning using raw biometric datasets which facilitate their development of being a robust and scalable multimodal system.

4. Multimodal Fusion: Multimodal fusion is the process of combining information from different modalities with weights to create a comprehensive biometric representation. This technique can occur at any processing stage such as feature extraction, representation learning, and decision making.

5. Deep Learning: A subfield within the means of machine learning that focuses on training neural networks using multiple layers to learn hierarchical representations of data. This method of learning has been given massive acknowledgment for its success in computer vision-related tasks that include face and iris recognition.

6. Limitation of a Unimodal Approach: Unimodal systems rely on one single source of biometric data which can lead to vulnerabilities. This approach is vulnerable to various environmental limitations that include illumination changes, occlusions, poses, expressions, and spoofing attacks. In terms of results, unimodal systems tend to suffer from a high false acceptance rate (FAR) and false rejection rate (FRR) which lead to security risks.

7. Strengths of a Multimodal Approach: Multimodal systems allow the combination of multiple biometric modalities to help enhance accuracy and robustness. By obtaining different biometric traits, multimodal systems can overcome certain limitations that an unimodal approach has. An example is while faces can be affected by illuminations and poses, iris patterns remain stable under different conditions. Furthermore, multimodal fusion techniques can integrate information through different modalities during various stages of processing hence improving the overall reliability of biometric identification systems. It is important to notice how multimodal systems are also more resistant to spoofing attacks as attackers would be required to spoof multiple biometric traits simultaneously which increases the complexity of an attack.

III. Dataset

The dataset used for this project with respect to testing and training the program was the VISA Face and Iris Multimodal Biometrics Database. This database was published online by the Visvesvaraya Technological University, Belagavi. It is a multimodal dataset that has captured images of face and iris data from the same individuals between the ages of 10 to 90 years old. The database used simple image acquisition techniques to gather the required biometric features from individuals in uncontrolled scenarios and variations [1]. The images contained a variety of samples where they were captured both indoors and outdoors.

The face images were captured using various cameras and scenarios hence resulting in reflections, shadows, occlusions, expressions, and pose variations being present among them. For the iris images, it was a similar outcome whereby variations, occlusions, and iris alignments were present. In the table below, we can view the overall demographical breakdown for the following database which includes the gender and age distribution among the 100 subjects [1].

TABLE I
VISA Database Demographic

Distribution Type	100 Subjects
Gender	Male: 65% Female: 35%
Age	10 - 25: 48% 26 - 40: 24% 41 - 55: 15% $\geq$ 56: 13%

The storage of the database was as such, the face images were stored in a "jpg" format whereas the iris images were stored in a ”bmp” format. Firstly, for the face dataset method of storage, folders were created for each subject where they were formatted using the following method, the first five characters indicated the subject ID (e.g S0000), followed by their gender (M = Male and F = Female), age (e.g 21), acquisition year (e.g 2024) and the image sequence number (e.g 001) [1]. Within each folder, we can see the images are laid out with the same format with only the image sequence number increasing depending on the number of images located within the folder itself.

Regarding the iris images, a similar format was followed but the folders only contained the subject ID, gender, and age. Within each folder, two additional folders are present and are labeled as L = Left, and R = Right indicating which are the left and right eye images that have been captured. Do note that, as with the face images mentioned above, the iris images are numbered using a sequence and they are incremental depending on the number of images captured for each eye.

A. Data Preparation Architecture/Pipeline

Figure 1 (below) shows the overall layout for both approaches used in this research, the K-Nearest Neighbor and Deep Learning Pipeline created for the dataset we have obtained. It starts by taking in an input image whereby it parses the data, then proceeds to preprocessing where it will crop, binarize, and prepare the data for the next step which is feature extraction where notable data will be obtained to then train the model we have implemented and created.

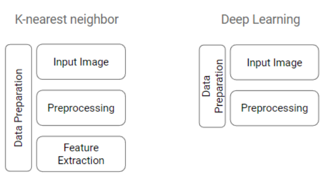

Fig. 1. Dataset Pipeline/Architecture

IV. Related Works

This research displays how using a multimodal approach is far more beneficial in comparison to an unimodal approach due to its robustness and level of security as it requires multiple types of biometric data that would be collected. The paper relates solely to the concept of face and iris recognition and how it is used as a form of biometric security. During our research for this project, we have come across various works and papers that have related to our approach for this research as well, the following are some notable works we have found in relation to our project:

1. Face and Iris Recognition Using Neural Network - A Review: This paper provided us with the initial idea of utilizing an RBNN structure for our neural network as the paper conducted numerous tests and analyses on different ANN structures and architectures. In the paper, their study provided a comprehensive examination of conventional deep learning-based techniques for facial and iris recognition. The methods that they assessed during the critical stages of image acquisition, preprocessing, feature extraction, and matching. In addition, databases for training or testing photo sets, performance measurements, techniques, algorithms, and procedures related to face recognition systems are investigated [2]. While other studies have sought to solve the shortcomings of the previously stated method, there are still certain advantages and disadvantages to the approaches that they have highlighted. Their paper discusses various deep-learning methods for biometric identity recognition, categorization, and identification. Furthermore, various biometric models—such as single-modal and multimodal datasets are employed for the recognition.

Used methodology	Recognition Accuacy of Model
“RCNN”	90.30%
“RINN”	90.60%
“MRC & MLP” NeuralNetwork	91.60%
PCA with ANN Face	95.45%
‘CNN”	85:1%
“DDFD”	91.79%
“RBNNs”	97.56%
“Gabor Wavelet Faces withANN”	93%
“WNN’	89.22%

Fig. 2. Artificial Neural Networks Scores Table [2]

2. Face and Iris-based Human Authentication using Deep Learning: This paper is one of the biggest cornerstones for our approach in this research. In the paper they highlighted three approaches to multi-modality fusion for biometrics, the first is image fusion, the second is score fusion and the the last one is feature fusion. We implemented a rudimentary version of score fusion due to the scope of the project and yielded impressive results. In the paper, they found that their suggested multimodal systems based on the image-level, feature-level, and score-level fusion achieved a high accuracy rate of 100% with the application of the two-fold cross-validation technique. Interestingly, the fused image they produced is the only image from which the features utilized by the multimodal system based on image-level fusion were derived. Furthermore, when they used only one sample for training and the rest of the database for testing, the score-level fusion multimodal proposed system achieved more than 99% accuracy on the SDUMLA-HMT multimodal database and 100% accuracy on the CASIA-ORL database [3]. Which is extremely impressive as compared to our 84.5% when we performed our rudimentary score fusion technique.

3. Deep Learning Based Face Recognition Method using Siamese Network: This work helped us to pinpoint and decide on the use of Siamese Neural Networks as it provides a lot of advantages through our use-case. One interesting point highlighted in the paper is how they suggested using Siamese networks to recognize faces instead of tagged face photos. They managed to accomplish this by using nearest neighbor counterparts and negative samples deliberately to create positive and negative pairs using an unsupervised technique [4]. This is an intuitive approach to the need to label all the images manually.

4. A Multimodal Biometric System for Iris and Face Traits Based on Hybrid Approaches and Score Level Fusion: This paper helped us to finally decide on utilizing a score-fusion method for our research. This provided us with a deep dive into the methodology and its benefits. The experimental results in their study show that their proposed technique performs better on several parameters, including precision, recall, accuracy, equal error rate, false acceptance rate, and false rejection rate. The accuracy rates of the face and iris biometric systems are 96.45% and 95.31%, respectively. The iris and face had equal error rates of 2.36% and 1.79%, respectively. Concurrently, their suggested multimodal biometric system achieves a significantly improved accuracy rate of 100% and a comparable error rate of as low as 0.26% [5]. Which is the result that we are trying to strive for in our case.

5. Quality Assessment-based Iris and Face Fusion Recognition with Dynamic Weight: This paper provided the idea of using weights and implementing them when performing a fusion between the two modalities. In the paper, they highlighted a plan that suggests a way to assess feature matching quality using the face and iris quality measures. Additionally, the plan suggests a dynamic weighted fusion approach to minimize the impact of subpar biometrics on the recognition outcomes and optimize the combination of these two biometrics' weights during the score-level fusion process [6]. As a result, dynamic weighted fusion recognition at the score level will be accomplished. One idea we planned to expand on this paper's idea is to train a Convolutional Neural Network to implement the dynamic implementation of weights when performing a fusion of the two modalities. We believe it would yield a much better result because it will remove human intervention and bias.

V. Method & Implementation

A. *Face Implementation*

1. Data Preparation - Parsing Input: The first step for face implementation we proceeded with is the processing of the dataset of face images that have been stored within the directory. Within this step and process, we would extract the metadata required such as the labels and IDs for each of the images, and then proceed to compile these data into a list.

```python
def parse_face_dataset() -> tuple[list[cv2.typing.MatLike], list[str]]:
    images: list[cv2.typing.MatLike] = []
    labels: list[str] = []

    # Iterate over each directory in the base directory
    for path in glob.iglob(base_directory + '/*'):
        # Extract the filename from the path
        folder_name = os.path.basename(path)

        # Parse the filename to extract the label
        match = re.search(r'(.*?)_2017_001', folder_name)
        if match:
            label = match.group(1)
        else:
            warnings.warn(f"No match found for filename: {folder_name}")
            continue

        for image_path in glob.iglob(path + '/*'):
            try:
                # Read the image as grayscale
                image = cv2.imread(image_path, cv2.IMREAD_GRAYSCALE)
```

Fig. 3. Parse Face Dataset Function

2. Data Preparation - Preprocessing: Next, we then proceed to process the face images found and detect these faces within each image by using the Haar Cascade Classifier. This would then continue to crop the images to whereby the face is grabbed more specifically and saves these cropped images into another directory. This is where we have also created a split to save the data into training and testing. For this paper, we have decided to split the data into 80% training and 20% tests respectively.

```python
face_cascade = cv2.CascadeClassifier(
    'Dependencies/haarcascade_frontalface_alt2.xml')

for face_image in face_images:
    (image, image_id, label) = face_image
    image_id += 1

    faces = face_cascade.detectMultiScale(image, 1.1, 4)
```

Fig. 4. Snipped of Face Cascade Classifier

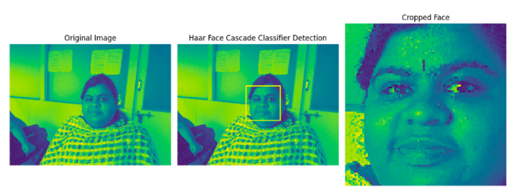

Fig. 5. Face Detection Process

3. Data Preparation - Feature Extraction: Once the faces have been detected and split, we continue to the feature extraction as this is how we can gather unique data for each individual for the network to identify each person through their distinct features. The program will detect each image stored previously and predicts facial features to extract important information by drawing lines and calculating the following:

- **Feature 1:** The distance between the eyes.
- **Feature 2:** The width and height of the nose.
- **Feature 3:** The width and height of the mouth.

Once that is completed, we move on to use an external dependency called Shape Predictor 68 to locate facial landmarks. This will plot 68 facial points on the images we have extracted

```python
> def calculate_eye_distance(shape):...
> def calculate_nose_shape(shape):...
> def calculate_lips_contour(shape):...
> def calculate_mouth_wrinkles(shape):...
> def draw_lines(image, shape):...
```

Fig. 6. Functions for Features Extracted

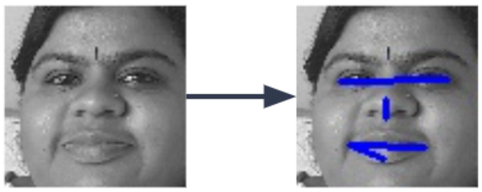

Fig. 7. Process of Line Drawing for respective features

and it will extract those coordinate points found and that have been drawn out on the faces. When that has been finished, the program will continue to flatten each set into a feature vector which is then saved into NumPy files to be used later on in the neural network.

```python
predictor = dlib.shape_predictor(
    'Dependencies/shape_predictor_68_face_landmarks.dat')

for i, d in enumerate(dets):
    shape = predictor(image, d)
    landmarks = [(shape.part(i).x, shape.part(i).y)
                 for i in range(68)]
    for (x, y) in landmarks:
        cv2.circle(image, (x, y), 1, (0, 0, 255), -1)
```

Fig. 8. Snippet of utilizing shape predictor dependency

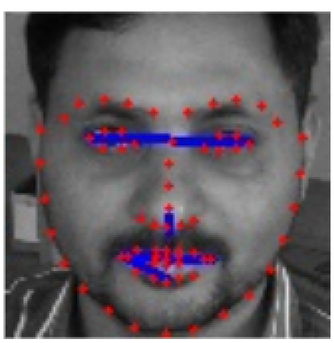

Fig. 9. Red points indicating landmark points on face

B. *Iris Implementation*

1. Data Preparation - Parsing Input: The first step for the iris implementation follows the same process as the face data, we would have to process the dataset of every eye image stored within the directory that includes, the images from both the left and right eye. We will proceed to extract the metadata once again such as the labels and IDs and compile them into another list.

```python
def parse_iris_dataset(keep_reflections: bool = False) -> tuple[list, list, list[str]]:
    left_images: list = []
    left_labels: list[str] = []
    right_images: list = []
    right_labels: list[str] = []

    for path in glob.iglob(base_directory + '/*'):
        folder_name = os.path.basename(path)

        label = folder_name

        # Process Left Eye
        for image_path in glob.iglob(path + '/L/*'):
            image = cv2.imread(image_path)

            image_hough_processed = process_hough(image_path, image, 50)
```

Fig. 10. Snippet of Iris Parsing Function

2. Data Preparation - Preprocessing: Once that has been done, we would continue to preprocess the eye images found and then conduct iris detection for each image using the Hough Circles algorithm. This would then crop the eye while keeping the iris image (more close up) intact, it then saves the new image into an output directory and will split the dataset into 80% training and 20% testing data once again. Initially, we attempted to do this using the Iris Cascade Classifier but decided to switch to Hough Circles instead.

```python
def process_hough(imagepath, image, radius):
    image = cv2.resize(image, (640, 480), interpolation = cv2.INTER_LINEAR)
    image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
    gray = cv2.medianBlur(image, 11)
    # edge = cv2.Canny(image, 100, 200)
    ret, _ = cv2.threshold(gray, 0, 255, cv2.THRESH_BINARY + cv2.THRESH_OTSU)

    circles = cv2.HoughCircles(gray, cv2.HOUGH_GRADIENT, 1,
                               50, param1=ret, param2=30, minRadius=20, maxRadius=100)
```

Fig. 11. Snippet of Hough Circles Function

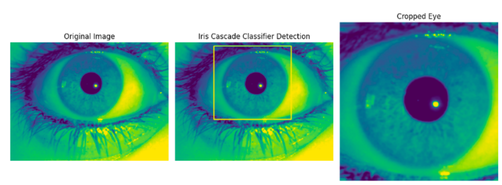

Fig. 12. Iris Detection Process

After that process has been completed, we will then get the images ready for feature extraction using two algorithms:

- **Hough Circles:** It would find circles (essentially the iris) in the image.
- **Daugman Rubber Sheet Model:** Then that image will be turned into a rectangular sheet image.

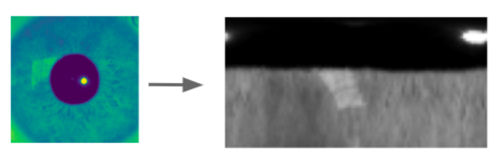

Fig. 13. Hough Circles to Daugman Rubber Sheet

3. Data Preparation - Feature Extraction: Continuing, we then used Discrete Wavelet Transform to extract important and needed features from the image. This would separate the data into different bands also known as "sheets". It also utilizes directional data whereby it gets the horizontal, vertical, and diagonal parts as well. The reasoning as to why we decided on this process is so that we can extract more feature data from the existing data, which in turn provides us with higher accuracy as well.

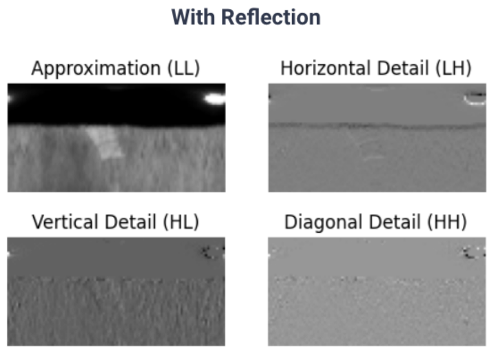

Fig. 14. Daugman Rubber Sheet Model With Reflections

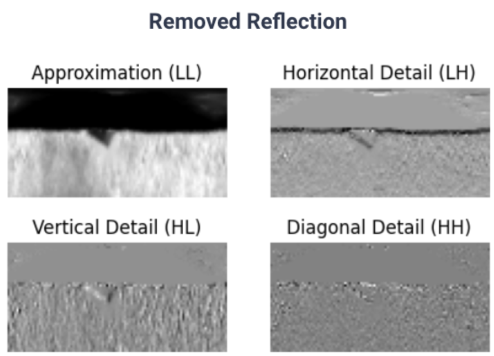

Fig. 15. Daugman Rubber Sheet Model With Reflections Removed

C. *Single Modal Implementation (Training and Test)*

In this section, we discuss the single model implementation for the **training** data.

- **K-Nearest Neighbor (KNN):** For the KNN approach, we trained two models using built-in scikit libraries for face and iris. We used a branch of KNN called the Ball-tree algorithm which is useful in terms of spatial indexing methods that are used in geometry and machine learning to organize multidimensional data. Moreover, this approach is faster than our neural network approach.
- **Deep Neural Network (Deep NN):** As for the deep neural network method, we trained three models using TensorFlow which are the L (Left) iris, R (Right) right, and face. In this method, we used the Siamese neural network which is an artificial neural network that uses the same weights to calculate output vectors from different input vectors. In addition, when using this method, we found it to be very slow when compared to the KNN approach.

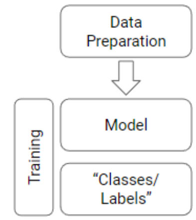

Fig. 16. Single Modal Training Flowchart

This section discusses the single model implementation for the **test** data.

- **KNN:** For the test data, the approach to using KNN was much simpler, we decided to just feed the model itself features from the iris and face model.
- **Deep NN:** For the neural network approach, we had to just feed the model images for the iris and face model.

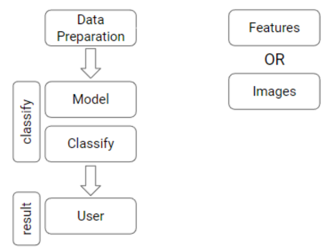

Fig. 17. Single Modal Testing Flowchart

D. *Score Fusion Approach*

The fusion score approach follows the same methodology as the single-modal but the difference is that this method would do a fusion on the scores it obtains at the end of execution. Fusion score combines match scores from different modalities to identify the specific individual. For our approach, we decided to also apply weights and then calculate the new score before it outputs the final score itself.

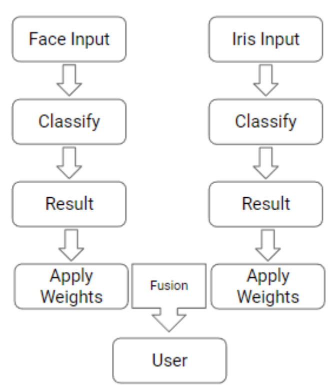

Fig. 18. Score Fusion Flowchart

E. Deep Neural Network Implementation: Siamese Neural Network

Before deciding on using the Siamese NN approach, we initially used a Radial Basis Neural Network (RBNN) but unfortunately, that yielded terrible results for what we were attempting to accomplish in this project. The Siamese NN uses few shots learning whereby is useful in our case as this method requires less data. Moreover, this approach essentially runs two neural networks simultaneously and will then calculate the similarities and compare the predictions to the true label itself.

```python
def create_siamese_network(input_shape, embedding_dim):
    input = layers.Input(shape=input_shape)
    x = layers.Conv2D(32, (3, 3), activation='relu')(input)
    x = layers.MaxPooling2D((2, 2))(x)
    x = layers.Conv2D(64, (3, 3), activation='relu')(x)
    x = layers.MaxPooling2D((2, 2))(x)
    x = layers.Conv2D(128, (3, 3), activation='relu')(x)
    x = layers.MaxPooling2D((2, 2))(x)
    x = layers.Conv2D(128, (3, 3), activation='relu')(x)
    x = layers.MaxPooling2D((2, 2))(x)
    x = layers.Flatten()(x)
    x = layers.Dense(512, activation='relu')(x)
    output = layers.Dense(embedding_dim)(x)
    return models.Model(input, output)
```

Fig. 19. Siamese Neural Network Function

```python
# Create Siamese network branches
base_network = create_siamese_network(input_shape, embedding_dim)
input_a = layers.Input(shape=input_shape)
input_b = layers.Input(shape=input_shape)

processed_a = base_network(input_a)
processed_b = base_network(input_b)

distance = tf.abs(processed_a - processed_b)

output = layers.Dense(1, activation='sigmoid')(distance)

model = models.Model([input_a, input_b], output)

model.compile(optimizer='adam', loss='binary_crossentropy',
              metrics=['accuracy'])
```

Fig. 20. Siamese Neural Network Branches

F. K-Nearest Neighbors (KNN) Implementation

Regarding the KNN, we created two models, one for train and test. For the train model, the KNN would plot points on a "graph" by using the features it had previously extracted. When it comes to the test model, it would look at the nearest "k-point" to it which is the feature. It will then assign that to the most common category.

```python
def train(
    x_train,
    y_train,
    n_neighbors: int | None = None,
    knn_algo: str = 'ball_tree',
    verbose: bool = False,
):
    if n_neighbors is None:
        n_neighbors = int(round(math.sqrt(len(x_train))))
        if verbose:
            print("Chose n_neighbors automatically:", n_neighbors)

    knn_clf = neighbors.KNeighborsClassifier(
        n_neighbors=n_neighbors,
        algorithm=knn_algo,
        weights='distance',
    )
    knn_clf.fit(x_train, y_train)

    return knn_clf
```

Fig. 21. Snippet from KNN Train Function

VI. Results & Discussion

A. Siamese Neural Network:

- **Single Modal (All dataset):**
  - **Face:** 83% (on average)
  - **Iris:** 93% (on average)
- **Multimodal (Targeted images):**
  The multimodal approach for Siamese utilizes the images directly from the iris and face to compare and match them accordingly.

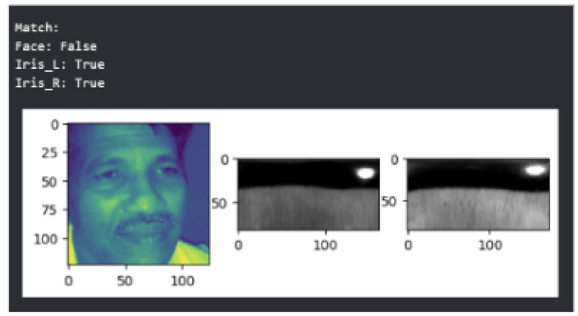

Fig. 22. Result Obtained From Siamese Multimodal Approach

We have also tested the dataset out on the following methods below to conduct a comparison with our approach:
  - **MobileNet:** 54% (on average)
  - **Custom RBNN:** 27% (on average)

For the Siamese NN, in terms of the single modal approach, for the face accuracy, we seem to get around 83% which is still lower when compared to the iris which obtained around 93%. When it comes to the multimodal approach, we used targeted images to identify if the face and iris datasets match with the specific individual. In the example above, we can see that the face was falsely detected whereas both iris images matched. We also attempted to try this on two other methods, MobileNet and a Custom RBNN modal but as seen from the results, they did not perform as expected and this was an interesting observation.

B. K-Nearest Neighbor (KNN):

- **Single Modal:**
  - Face: 73% (on average)
  - Iris: 96% (on average)
- **Multimodal:**
  - Balanced Weight: 84.5% (on average)
  - Skewed Iris (30:70): 89.1% (on average)
  - Skewed Face (70:30): 79.9% (on average)

From the KNN approach, we can notice that the single modal method for the face did not produce a high accuracy when compared to the iris approach. In terms of multimodal, we can see that if we apply a balanced weight on both the iris and face, we obtain a result of around 84.5%. But if we were to apply more weight to the iris portion, it seems to perform much better than if we skew the face weights more.

VII. Conclusion & Thoughts

To conclude, we can notice that the results obtained can be enhanced through the employment of weights to the models. This can also compensate for the less ideal results that are obtained from the single-modal approach. Concerning the dataset, the dataset we used for this paper was a little inconsistent for some of the classes as some classes lacked images and were inconsistent in the number of images found for each subject and folder. Due to our idea and approach being fairly new to the industry, it was a challenge to find the required dataset we needed to enhance and improve our project overall. Lastly, when the data sample is insufficient to begin with, we found that using one-shot/few-shot learning architectures was able to greatly aid us in improving our overall results for this research. This project has given great insight into the use of a multimodal approach for biometric security and we hope that in the future, we will be able to expand on this idea by improving and experimenting with more methods and techniques.

References

[1] S. A. Angadi and V. C. Kagawade, "Face and iris wavelet feature fusion through canonical correlation analysis for person identification," Dec. 2018.

[2] S. Rani, N. Mohan, and P. Kaushal, "Face and iris recognition using neural network: A review," pp. 1–6, 2023.

[3] B. Muthukumaran, L. Harshavarthanan, S. Dhyaneshwar, and M. Z. Sharief, "Face and iris based human authentication using deep learning," pp. 841–846, 2023.

[4] E. Solomon, A. Woubie, and E. S. Emiru, "Deep learning based face recognition method using siamese network," 2023. [Online]. Available: https://arxiv.org/abs/2312.14001

[5] O. Najah Kadhim, M. Hasan Abdulameer, and Y. Mahdi Hadi Al-Mayali, "A multimodal biometric system for iris and face traits based on hybrid approaches and score level fusion," *BIO Web of Conferences*, vol. 97, p. 00016, 2024. [Online]. Available: http://dx.doi.org/10.1051/bioconf/20249700016

[6] K. Xiao, Y. Tian, Y. Lu, Y. Lai, and X. Wang, "Quality assessment-based iris and face fusion recognition with dynamic weight," *The Visual Computer*, vol. 38, no. 5, p. 1631–1643, Mar. 2021. [Online]. Available: http://dx.doi.org/10.1007/s00371-021-02093-7