A REPORT

ON

FACE RECOGNITION USING SIAMESE NEURAL NETWORK

By

Name of the student	Registration No.
1) Om Sai Vasireddy	AP21110011282
2) Uma Maheswar Reddy Nelli	AP21110011305
3) Sai Charan Teja Janaki	AP21110011314
4) Vikram Muchumarri	AP21110011337

Prepared in the partial fulfillment of the
Undergraduate Research Opportunities Project (UROP)

AT
SRM University, AP
Neerukonda, Mangalagiri , Guntur
Andhra Pradesh – 522 240

Under the Guidance of Dr. Mudassir Rafi
November 2023

Certificate

Date: 26-11-2023

This is to certify that the work present in this Project entitled **"Face recognition using Siamese neural network"** has been carried out by **'Om Sai Vasireddy', 'Uma Maheswar Reddy Nelli', 'Sai Charan Teja Janaki', 'Vikram Muchumarri'** under my supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology in School of Engineering and Sciences.

Supervisor
Dr. Mudassir Rafi,
Assistant Professor,
Computer Science Department.

Acknowledgments

We would like to express our profound gratitude to Dr.Niraj Upadhayaya, of the CSE department, and Dr.Mudassir Rafi for their contributions to the completion of my project titled "**Face recognition using Siamese Neural Network**".

We would like to express our special thanks to our mentor Dr.Mudassir Rafi for the time and efforts, he provided throughout the year. Your useful advice and suggestions were helpful to us during the project's completion. In this aspect, We are eternally grateful to you.

We would like to take this opportunity to express our gratitude to all of our group members. The project would not have been successful without their cooperation and input.

Om Sai Vasireddy

Uma Maheswar Reddy Nelli

Sai Charan Teja Janaki

Vikram Muchumarri

Table of Contents

ABSTRACT

Today the field of Face Recognition and its applications are experiencing rapid development. Facial Recognition plays a vital role in identifying individuals, even setting apart identical twins. It has become widely adopted in contemporary applications for authentications such as phone unlocking, and criminal identification. Identifying and extracting relevant information from the input data are crucial for training a machine learning model. The One-shot learning method, where we have only given one example of each new class and must make accurate predictions. To automatically prioritize inputs we use a technique for training *Siamese neural networks*. Once we fine-tune the network, we can use its strong features to make predictions not only on new data but also on completely new categories that it hasn't seen before. Convolutional Neural Network (CNN) is one of the deep learning techniques used in face recognition. Using this convolutional architecture, we get better results that do better than other models, especially in tasks where we only have one example for each category.
Keywords—Face Recognition; Machine Learning; Deep Learning; Siamese Neural Network; Convolutional Neural Network;

SRM UNIVERSITY AP
Andhra Pradesh

Abbreviations

- CNN - Convolutional Neural Network
- SNN - Siamese Neural Network
- SVM - Support Vector Machines
- ML - Machine Learning
- DL - Deep Learning
- ReLU - Rectified Linear Unit
- RBF - Radial Basis Function
- RGB - Red, Green, Blue
- HBPL - Hierarchical Bayesian Program Learning

List of Figures

List of Tables

List of Formulae

- $L(x_1^{(i)}, x_2^{(i)}) = y(x_1^{(i)}, x_2^{(i)})\log p(x_1^{(i)}, x_2^{(i)}) + (1 - y(x_1^{(i)}, x_2^{(i)}))\log(1 - p(x_1^{(i)}, x_2^{(i)})) + \lambda|w|^2$

1. Introduction

1.1 The Importance of Face Recognition

In today's technological landscape, face recognition stands as a pivotal tool across various domains. Utilizing advanced methods like deep learning, it allows for the identification and distinction of individuals based on facial features. Its applications range from security systems to personalized interactions, presenting far-reaching implications [1].

1.2 Purpose of the Report

This report aims to explore the realm of face recognition technology, specifically focusing on refining the performance of convolutional neural networks (CNNs) through the integration of Siamese networks. The primary objective is to enhance accuracy, generalization, and inference speed by employing state-of-the-art optimization techniques such as rectified Adam [2].

1.3 Background of Face Recognition

The evolution of face recognition heavily relies on deep learning methodologies, particularly CNNs. These networks, comprising convolutional, pooling, and fully connected layers, have become instrumental in processing visual information for facial recognition systems [3].

1.4 Issue at Hand

While traditional methods have proven effective, there's a growing need for improved efficiency and accuracy, especially concerning one-shot learning and rapid verifications. This research addresses these challenges by leveraging Siamese networks [4].

1.5 Objectives of the Research

This study aims to develop a robust face recognition model using Siamese neural networks. The primary focus lies in creating a model capable of accurate classifications and verifications based on minimal examples, thereby enhancing adaptability and performance [5].

1.6 Relevance of the Study

In the era of rapid advancements in deep learning, face recognition has become an indispensable application. The utilization of Siamese networks presents an innovative approach, particularly in the context of one-shot learning, catering to diverse practical applications [6].

1.7 Data Collection Methodology

For this research, the dataset employed originates from the 'Labelled Faces in the Wild' repository, providing a diverse array of facial images for training and validation [7].

1.8 Proposed Solution Approach

The methodology hinges on Siamese neural networks to measure similarity and dissimilarity between facial features. Leveraging the L1 distance metric, these networks facilitate accurate verifications based on minimal exemplars, enhancing the system's robustness [8].

1.9 Research Goals and Targets

The primary objective is to advance face recognition capabilities by implementing Siamese networks for efficient one-shot learning. Specific goals include optimizing accuracy, enabling swift verifications, and contributing novel insights to the field of facial recognition technology.

1.10 Patterns Observed in Data

The observations will primarily revolve around distinctive patterns and features in facial images crucial for accurate recognition. The intent is to develop a system proficient in discerning unique characteristics, thereby contributing to the understanding of facial recognition methodologies.

1.11 Insights from the Research

Insights derived from this research will deepen the understanding of Siamese network architectures for face recognition. By enhancing one-shot learning capabilities, the system aspires to offer adaptable and robust solutions for various real-world applications.

2. Methodology

Machine Learning and Deep learning models can be employed in classification of face images collected from various datasets. The models that we have used in our research are CNN and SNN.

2.1 Models

a. **Convolutional Neural Network (CNN)**

CNNs, designed for handling grid-like data such as images, form the cornerstone of our classification methodology.

Convolution Layers: These layers operate by applying filters (kernels) to input images, detecting distinctive patterns or features within localized regions, enabling the network to recognize complex structures.

Pooling Layers: Serving to condense information, pooling layers down sample feature maps, aiding computational efficiency and promoting translational invariance in learned features.

Dense Layers: Positioned at the end of the network, these fully connected layers amalgamate extracted features, enabling high-level abstractions for final classification.

These layers in concert allow the network to progressively abstract and learn intricate features within images, crucial for accurate classification tasks.

b. **Siamese Neural Network**

The Siamese Neural Network (SNN) architecture is engineered for tasks involving limited labeled data or one-shot learning scenarios. It revolves around comparing and recognizing the similarity or dissimilarity between pairs of inputs.

SNN in One-Shot Learning: One-shot learning refers to a paradigm where models make predictions or classifications based on a single example per class. The Siamese architecture excels in this scenario by learning and encoding similarity relationships between individual pairs of images. It discerns between different classes using minimal training samples, crucial in sparse datasets.

Learning Representations: SNNs learn representations or embeddings for input pairs, capturing underlying similarities or differences between images. These representations encode essential visual features enabling meaningful comparisons in a feature space where similarity metrics are computed.

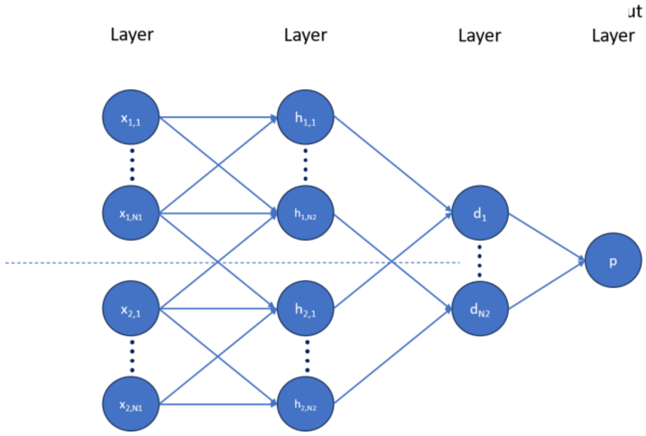
Contrastive Loss Function: Central to SNN training is the contrastive loss function. This guides the network to learn embeddings that minimize the distance between similar image pairs while maximizing the distance between dissimilar pairs. The loss function optimizes the network parameters for effective discrimination based on learned embeddings.

L1 Distance Layer: The L1 distance layer computes the Manhattan distance between the learned embeddings of two input images. Mathematically, the L1 distance between vectors

$$L(x_1^{(i)}, x_2^{(i)}) = y(x_1^{(i)}, x_2^{(i)})\log p(x_1^{(i)}, x_2^{(i)}) + (1 - y(x_1^{(i)}, x_2^{(i)}))\log(1 - p(x_1^{(i)}, x_2^{(i)})) + \lambda|w|^2$$

This layer facilitates the calculation of the absolute differences between corresponding

Figure 1.Flow chart of a general Siamese model network



elements of the embeddings, providing a direct measure of dissimilarity or similarity between pairs of images.

2.2 Literature Review

One-shot learning, an early endeavor in machine learning, aims to achieve accurate predictions or classifications with minimal training data [9]. Pioneering work in the early 2000s by Li Fei-Fei et al. focused on one-shot learning for character recognition, employing Hierarchical Bayesian Program Learning (HBPL). HBPL's approach involved decomposing images into smaller components to discern the structural significance of these elements. However, the computational complexities associated with HBPL posed significant challenges in inference.

In 2009, Maas and Kemp made significant strides by enhancing Bayesian networks to predict attributes within Ellis Island passenger data, demonstrating the versatility of Bayesian approaches beyond face recognition. Lim's 2012 work primarily addressed the practicality of handling limited datasets for specific classes, particularly employing Convolutional Neural Networks (CNNs) for face recognition [10].

The application of CNNs has been pivotal in various face recognition problems [11] (Li & Hua 2015; Parchami et al. 2017). CNNs, a type of feedforward neural network with convolutional computation and deep learning architecture, extract color, shape, and texture features from image datasets using RGB-level co-occurrence matrices. These features undergo selection tasks facilitated by a series of convolutional layers, employing diverse filters of varying sizes. To optimize efficiency, the quantity of convolutional filters is typically set as multiples of 16.

The architecture of CNNs for face recognition integrates Rectified Linear Unit (ReLU) activation functions post convolutional layers. This incorporation introduces non-linearity into the model, enabling the capture of intricate relationships within the data [12]. Further optimization of the model is achieved through the Adam optimizer.

Concluding with a 4096 fully connected layer, the network computes the L1 component distance between vectors. This computation of the L1 distance is pivotal, enabling effective comparison and evaluation of similarity among different facial feature representations, a crucial aspect in face recognition models[13].

The evolution of one-shot learning methodologies, particularly within face recognition leveraging CNN architectures, highlights the significance of feature extraction, architectural design, and the pivotal role of the L1 distance computation in evaluating and comparing facial feature similarities.

2.3 Architecture

The architectural design for face recognition with Convolutional Neural Networks (CNNs) and Siamese Neural Networks (SNNs) forms the cornerstone of our approach.

Datasets Information Recap
Our dataset, a compilation from Labeled Faces in the Wild Home and self-generated images via OpenCV, consists primarily of three image categories: anchor, negative, and positive. Anchors are input images, paired with either negative (from the dataset) or positive (user-generated) images for training purposes.

Preprocessing
Resizing images to a uniform 100x100 pixel dimension standardized the dataset for both training and testing. Preceding the data split, extensive augmentation fortified the model, ensuring robustness and easing the training process. Augmentation merged anchor images with negatives (labeled 0) and positives (labeled 1), enabling parallel image feeds into similar networks to compute their differences at the output layers.

Siamese Neural Network (SNN) Implementation

Our robust and efficient model incorporates Siamese networks, integrating CNNs. The architecture encompasses convolution layers, activation functions, feature maps, loss functions, and the crucial Euclidean distance computation. This model undergoes rigorous training using augmented data to discern various image structures and shapes, subsequently classifying them based on learned representations.

Network Components

Activation functions, such as rectified linear units (ReLU) in initial layers and sigmoidal units in subsequent layers, optimize feature map outputs. The final convolutional layer's units are flattened into a single vector, followed by a fully-connected layer. An additional layer computes the induced distance metric between Siamese twins, culminating in a single output layer employing sigmoid as its activation function.

Objective and Optimization

A cross-entropy objective is imposed on our binary classifier, reinforcing the discrimination between identical and distinct image pairs within a mini-batch. Standard backpropagation is employed, leveraging tied weights across twin networks, thus facilitating additive gradient actions.

Beyond Optimization

Post-optimization, the discerning features learned through one-shot learning are evaluated for their discriminative potential, amplifying the understanding of the network's acquired representations.

Additional Considerations

The architecture, while founded on established methodologies, maintains an adaptable framework for future iterations. Potential enhancements may involve nuanced activation functions, novel loss functions, or alternative network configurations to further bolster performance.

2.4 Implementation

Data Collection and Preprocessing

The dataset compilation involved the acquisition of facial images classified into anchor, negative, and positive categories. Images were sourced from the Labeled Faces in the Wild Home dataset, supplemented by self-generated images using OpenCV. A standardized resolution of 100x100 pixels was established for uniformity. Augmentation techniques, such as rotation, shifting, and flipping, were applied to bolster the dataset's variability. Labels were encoded to facilitate multi-class classification.

TABLE I. DETAILS OF THE DATASETS

Datasets	No of Images
Anchor Images	804
Negative Images	13,233
Positive Images	647

Dataset Processing and Labeling

The dataset was segregated into three distinct categories: anchors, positive pairs, and negative pairs. Anchors represented individual images, while positive pairs consisted of images from the same subject, and negative pairs contained images from different subjects. This arrangement allowed for supervised metric-based learning, essential for Siamese Neural Network (SNN) training.

Preprocessing - Scale and Resize

```
In [9]: def preprocess(file_path):

            byte_img = tf.io.read_file(file_path)
            img = tf.io.decode_jpeg(byte_img)

            img = tf.image.resize(img, (100,100))
            img = img / 255.0

            return img
```

Figure 2.Data Preprocessing function

Model Architecture: Convolutional Neural Network (CNN)

The CNN architecture was structured with convolutional layers followed by max-pooling for spatial reduction. Convolution layers integrated Leaky ReLU activations to capture feature representations. The model culminated in fully connected layers to facilitate classification tasks. A detailed summary of the CNN architecture, including layer configurations, was devised to harness the facial feature extraction capabilities of the CNN.

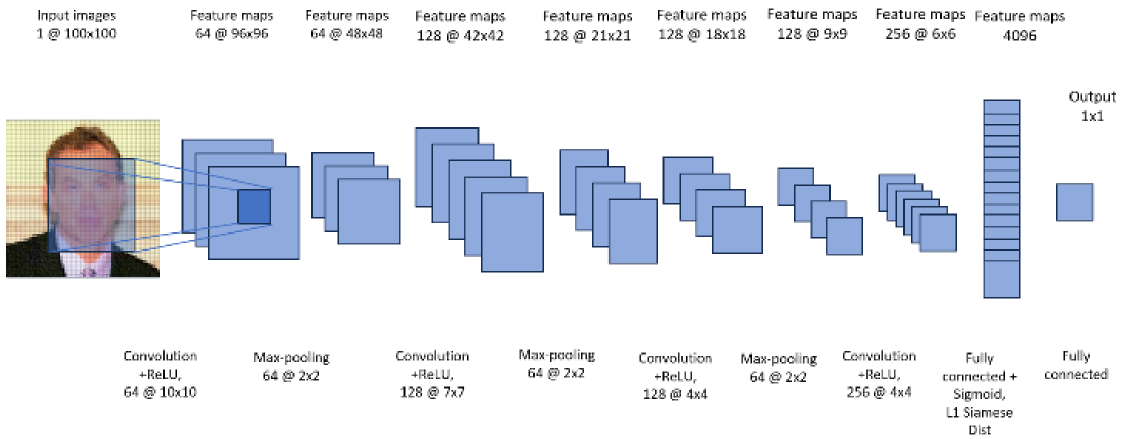


Figure 3. Convolutional Neural Network of our Model

Feature Extraction for Siamese Neural Network (SNN)

Features pivotal for SNN learning were extracted from intermediate layers of the pre-trained CNN model. A specific dense layer 'my_dense' functioned as a feature extraction point, aiding in the generation of representations conducive for one-shot learning without necessitating retraining.

Siamese Neural Network (SNN) Model Development

A dedicated SNN model was constructed, employing a customized layer 'L1Dist' to compute the L1 distance between embeddings extracted from the CNN. The Siamese network incorporated distinct layers for feature extraction, L1 distance calculation, and classification, culminating in a sigmoidal output layer for distance metric computation.

Training and Evaluation

The SNN model underwent training using an Adam optimizer and binary cross-entropy loss. The training process was iterated over epochs, and the model was evaluated using precision, recall, and confusion matrix analysis. Visualization techniques, such as Precision-Recall curves, provided insights into model performance, aiding in the evaluation and comparison of the trained model

TABLE II. EVALUATED VALUES OF ONE BATCH OF IMAGES AFTER TRAINING THE MODELS

VALIDATION IMAGE	LABEL	PREDICTED VALUES
Negative	0	1.0303144e-06
Negative	0	8.2426457e-05
Positive	1	9.9999893e-01
Positive	1	1.000000e+00
Negative	0	1.7540457e-05
Negative	0	1.6903167e-07
Negative	0	2.1551946e-06
Positive	1	9.8448521e-01
Negative	0	1.1103551e-05
Positive	1	9.9999082e-01
Positive	1	1.000000e+00
Positive	1	9.9933481e-01
Positive	1	1.000000e+00
Negative	0	1.4063612e-06
Positive	1	7.2294313e-01
Negative	0	5.8970704e-07

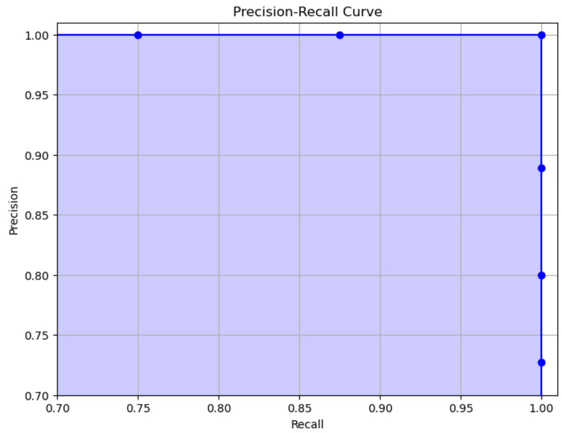


Figure 2 Precision-Recall Curve

Model Application and Verification

The trained SNN model was saved for future use and employed in a real-time verification setup. Utilizing an integrated webcam, the system captured input facial images, triggering a verification process. The system compared the captured image with stored verification images to authenticate and determine the subject's identity, presenting a verification result.

3. Discussion

Dataset Distribution Insights

Upon examination of our face recognition dataset, a graphical representation was generated using Matplotlib and Seaborn libraries to elucidate the distribution of classes. Dynamically calculating the image count per class using Python's OS library, we visualized the class distribution via a bar graph. This graphical depiction illustrates the disparity in the number of images across different classes, providing foundational insights into the dataset's composition. Understanding such distributions is pivotal for subsequent analysis and model training, ensuring an informed approach to handling class imbalances.

Visual Representation of Dataset Samples

Utilizing Matplotlib, we crafted a visual grid showcasing a subset of images from various classes within the dataset. Each column in this grid represents a distinct class, offering a visual glimpse into the diversity of facial images encompassed in our dataset. These visualizations serve as an introductory overview, aiding in understanding the variety and characteristics encapsulated within different class labels.

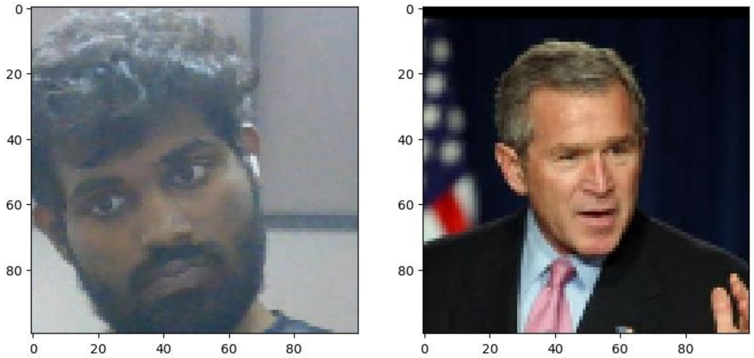


Figure 5. A sample dataset of an input and a negative image

Addressing Class Imbalance

In our face recognition dataset, notable variations exist in the distribution of images across different classes. Imbalances among classes, where certain classes exhibit significantly fewer instances compared to others, pose challenges during model training. For instance, observations revealed discrepancies in the number of images across specific facial features or expressions. Such imbalances can potentially lead to biased model predictions favoring overrepresented classes. Strategies to address these imbalances, such as data augmentation, class weighting, or sampling techniques, must be considered to ensure that the model learns robust representations across all facial attributes, thus bolstering its ability to generalize effectively.

Facial Recognition Techniques

We delved into various techniques, leveraging OpenCV's capabilities for facial recognition tasks. Image preprocessing, encompassing resizing and normalization, laid the groundwork for subsequent operations. Techniques such as Canny edge detection and contour extraction were employed to identify and visualize facial features and structures. The comparative analysis between the original images, Canny edges, and extracted contours showcased the efficacy of our techniques in detecting facial structures, forming a fundamental component of our facial recognition approach.

4. Conclusion

The unheard-of significance of Siamese networks in packages which includes facial recognition lies of their incredible efficacy. In contrast to conventional CNN-based deep getting to know, these networks demand fewer photograph inputs, considerably slashing training periods while effects computing similarity scores among photograph pairs—a important gain within the expansive subject of pc imaginative and prescient. Our exploration embarked with the strategic use of convolutional Siamese neural networks, validating their navigational prowess and delving into the intricacies of 1-shot type. Leveraging OpenCV, we meticulously crafted a based dataset, validating and refining the formidable abilities inherent in Siamese neural networks. Progressing to actual-time datasets, a pivotal shift befell, notably enhancing the version's performance metrics. Diverse metric assimilation no longer most effective scrutinized the version's overall performance but also unveiled its latent ability throughout diverse domain names, significantly in image category. Looking ahead, our focus sharpens on optimizing operations, diving into layers to minimize error margins amongst picture pairings. Augmenting with extra convolutional layers and leveraging optimization algorithms guarantees delicate version efficacy and the recognition of facial structures throughout diverse situations. This transformative journey anticipates the improvement of more strong popularity structures, no longer handiest in facial popularity but also in broader image-primarily based category, signaling a new generation of precision in system gaining knowledge of and computer vision technologies.

5. Future Work

Looking ahead, our number one consciousness is on pleasant tuning the version's operations for finest overall performance. This involves a meticulous exploration of various layers, meticulously assessing the space mistakes within diverse combinations of anchor, high quality, and poor pics. The aim right here is to minimize those errors, making sure unique results for each potential pairing. This elaborate system lets in for the extraction of nuanced image features, empowering the version to evolve and supply an increasing number of correct results. Expanding the community with extra convolutional layers and utilizing max-pooling to refine output distances among check and validation pictures emerges as a promising street for improvement. Employing the Adam Optimizer to constantly reduce errors for the duration of model schooling holds the important thing to predicted overall performance enhancements. Moreover, strategically augmenting these layers holds the capability to beautify the network's capacity to figure facial structures across diverse contexts, thereby amplifying its proficiency in diverse eventualities. These ongoing optimizations are geared toward fortifying the model's effectiveness, paving the manner for greater resilient and precise recognition systems within the foreseeable destiny.

6. REFERENCES

[1] The Handbook of Face Recognition by Stan Z. Li and Anil K. Jain (2011)

[2] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097–1105, 2012.

[3] C. Zhang, and Z. Zhang, “Improving multiview face detection with multi-task deep convolutional neural networks,” IEEE Winter Conference on Applications of Computer Vision, 2014, pp. 1036-1041.

[4] Chu, Z., Zhang, M., & Gong, X. (2021). Limited data face recognition: From theory to application. Springer.

[5] G. Koch, R. Zemel and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition", Proceedings of International Conference on Machine Learning, vol. 2, 2015.

[6] Chandhok, S. “Face Recognition with Siamese Networks, Keras, and TensorFlow,” PyImageSearch, P. Chugh, A. R. Gosthipaty, S. Huot, K. Kidriavsteva, R. Raha, and A. Thanki, eds., 2023.

[7] Niveditha M, Ramachandra Hebbar, Shashank N, Pooja R, Shima Ramesh, Prasad Bhat N, MVJ Engineering college, Karnataka, India, DOI: 10.1109/ICDI3C.2018.00017.

[8] Weinberger, K.Q.; Blitzer, J.; Saul, L.K. Distance metric learning for large margin nearest neighbor classification. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 10 December 2005; pp. 1473–1480.

[9] Palatucci, Mark, Pomerleau, Dean, Hinton, Geoffrey E, and Mitchell, Tom M. Zero-shot learning with semantic output codes. In Advances in neural information processing systems, pp. 1410–1418, 2009.

[10] Fe-Fei, Li, Fergus, Robert, and Perona, Pietro. A bayesian approach to unsupervised one-shot learning of object categories. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, pp. 1134– 1141. IEEE.

[11] K. Yan, S. Huang, Y. Song, W. Liu and N. Fan, "Face recognition based on convolution neural network," 2017 36th Chinese ControlConference (CCC), Dalian, 2017, pp. 4077-4081, doi:10.23919/ChiCC.2017.8027997

[12] Face Recognition Based on Convolution Neural Network. Kewen Yan,Shaohui Huang, Yaoxian Song1, Wei Liu1, Neng Fan.Murk Chohan, Adil Khan, Saif Hassan Katper, and Muhammad Saleem Mahar, “Plant Disease Detection using Deep Learning”, 2020, Begum Nusrat Bhutto Women University Sukkur, Sukkur Institute of Business Administration.

[13] IntechOpen. (2021). Face Recognition Using Siamese Networks with L1 Distance Metric. IntechOpen.