MASKED FACE RECOGNITION USING SUPPORT VECTOR MACHINE

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

M.Sc COMPUTER SCIENCE

ABINESH B

Reg.No: PCS052101

UNDER THE GUIDANCE OF
Dr. ADITHYA V

DEPARTMENT OF COMPUTER SCIENCE

SCHOOL OF PHYSICAL SCIENCES

CENTRAL UNIVERSITY OF KERALA

TEJASWINI HILLS PERIYE, KASARAGOD - 671320,

KERALA, INDIA AUGUST 2023

Department of Computer Science, Central University of Kerala
Tejaswini Hills, Periye - 671320, Kasaragod

CERTIFICATE

This is to certify that the thesis entitled, **"MASKED FACE RECOGNITION USING SUPPORT VECTOR MACHINE"** submitted by **ABINESH B (REG. NO: PCS052101** in partial fulfillment of the requirements for the award of M.Sc in Computer Science at the Central University of Kerala, is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the report has not been submitted to any other University Institute for the award of any degree.

DATE :

Dr. ADITHYA V
Assistant Professor
Department of Computer Science
Central University of Kerala
Kasaragod, Kerala - 671320

Department of Computer Science, Central University of Kerala
Tejaswini Hills, Periye - 671320, Kasaragod

CERTIFICATE

This is to certify that the thesis entitled, **"MASKED FACE RECOGNITION USING SUPPORT VECTOR MACHINE"** is a bonafide work carried out by **ABINESH B (REG. NO: PCS052101)** in partial fulfillment of the requirements for the award of M.Sc in Computer Science at the Central University of Kerala, Kasaragod, during the academic year **2021-2023**.

The work is satisfactory to award Master's Degree in Computer Science.

DATE :

1. EXTERNAL EXAMINER:

2.INTERNAL EXAMINER:

Dr. JAYA SUDHA J. S
Professor
Head Of the Department
Department of Computer Science
Central University of Kerala

DECLARATION

I, ABINESH B, Reg No: PCS052101, student of Fourth Semester M.Sc Computer Science, Central University of Kerala, do hereby declare that the report entitled, **"MASKED FACE RECOGNITION USING SUPPORT VECTOR MACHINE"**, submitted to the Department of Computer Science is an original record of studies and bonafide work carried out by me from APRIL 2023 to AUGUST 2023.

DATE: **ABINESH B**

PLACE: **PCS052101**

ACKNOWLEDGEMENT

If words are considered as a token of approval and symbol of acknowledge, then words play a heralding role in expressing our gratitude. I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I wish to express my deepest sense of gratitude to my guide Dr. ADITHYA V, Assistant Professor, Department of Computer science, Central University of Kerala. I cannot say thank you enough for her general support, constant encouragement and healthy criticism have been of valuable help to me throughout this project. I feel motivated and encouraged every time I attend her meeting. Without her encouragement and guidance, this project would not have materialized.

I also convey my heartful thanks to the Head of the Department, Dr. JAYA SUDHA J. S., for providing access to the laboratory and research facilities. Without her precious support it would not be possible to conduct this research.

I also thank our Teachers, Staffs, and Friends of the Department of Computer Science for sharing their knowledge and suggestions with me.

ABINESH B

PCS052101

Abstract

Face recognition technology has found widespread applications in security, surveillance, and user authentication. However, the COVID-19 pandemic introduced the challenge of identifying individuals wearing masks, as traditional face recognition systems rely on full facial features for accurate identification. This paper presents a novel approach for Masked Face Recognition (MFR) that combines deep learning with Support Vector Classification (SVC) to address this issue. By leveraging FaceNet, a deep learning model for face recognition, the proposed method extracts facial embeddings that encapsulate crucial facial information even when individuals are wearing masks. The method involves training on both masked and unmasked images of individuals, enabling the model to learn distinctive features from both scenarios. The trained FaceNet embeddings serve as input to the SVC, which learns to differentiate between individuals, considering both masked and unmasked scenarios. This fusion of deep learning and SVC results in an efficient and accurate masked face recognition system. The approach is evaluated on the COMASK20 dataset [1] and a self-collected dataset and got accuracies of 99.30% and 97.18% respectively.
keywords:Face Recognition, FaceNet, SVC, MFR

Table of contents

List of tables

Chapter 1

INTRODUCTION

1.1 Introduction

Recent years have witnessed the widespread integration of face recognition technology into various domains, transforming security, surveillance, and user authentication systems. However, the unexpected global adoption of masks due to events like the COVID-19 pandemic has introduced formidable hurdles for traditional face recognition methods. These systems, designed to identify individuals based on complete facial features, encounter challenges when faced with mask-wearing scenarios, resulting in diminished accuracy and reliability.

This research presents an innovative solution to overcome the limitations of conventional face recognition in the presence of masks. We introduce a novel approach for masked face recognition that synergizes deep learning and the classification of the Support Vector Classifier (SVC) [2]. Furthermore, we make a unique contribution by creating a self-collected data set specifically tailored to capture facial information in diverse mask-wearing variations.

At the core of our methodology lies the utilization of Face-Net[3], a cutting-edge deep learning model renowned for its capacity to discern discriminative features from facial images and map them into high-dimensional embeddings. By training Face-Net on our self-collected data set, encompassing a wide array of individuals donning various mask types, we empower the model to generate robust embed-dings that encapsulate vital facial attributes, transcending the mask-obscured regions.

A significant contribution of our study is the creation of this self-collected data set. Unlike conventional data sets that predominantly focus on unmasked faces, ours mirrors real-world scenarios where masks of varying types, colors, and coverage are worn. Training on this diverse data set equips the Face-Net model to adeptly recognize faces under different mask-related circumstances, ensuring accuracy and reliability.

The pivotal role played by the Support Vector Classifier (SVC) in our masked face recognition system is underscored. Leveraging the knowledge encapsulated in the Face-Net embeddings, the SVC excels in classification, distinguishing between individuals in both masked and unmasked scenarios. This amalgamation of deep learning and classification augments system accuracy, making it a practical choice for real-world applications.

To validate our approach, we conduct a comprehensive comparative analysis. Our masked face recognition system is benchmarked against traditional, mask-agnostic face recognition methods. Performance evaluations are conducted across public masked face data sets, conventional face recognition data sets, and our self-collected data set. This exhaustive examination underscores the superiority of our approach in mask-inclusive conditions, emphasizing the importance of accommodating mask-related variations for reliable face recognition.

Results from our experiments showcase the supremacy of our masked face recognition system on masked face data sets while maintaining competitive performance on traditional data sets. The incorporation of our self-collected data set further bolsters the credibility and applicability of our findings, underscoring the practicality of our approach in real-world mask-wearing scenarios.

In summary, this research advances masked face recognition technology by proposing an accurate and effective solution to the challenges introduced by mask-wearing. Our integrated approach, blending Face-Net with SVC, alongside the creation of the self-collected data set, holds promise across applications like security, surveillance, and user authentication, amidst the backdrop of prevalent mask-wearing practices

1.2 Organization of Thesis

The thesis is organized as follows. Chapter 2 provides the literature reviews in the area of Masked face recognition with Face-Net and SVC. Chapter 3 discusses the methodology involved in proposed research. Including collection, feature extraction using Face-Net & classification using Support Vector Classification. Chapter 4 describes the experimental analysis and results. Chapter 5 concludes the work and gives an idea about the future works in the research.

Chapter 2

LITERATURE REVIEW

The face recognition systems encountered many challenges after the pandemic which introduces the need of face mask. Numerous innovative solutions have been developed to achieve precise individual identification, even under conditions where subjects are wearing masks—items that can obscure substantial facial areas, including the nose. The occlusion issue emerges as a primary limitation within the practical implementation of two-dimensional (2D) face recognition methodologies. As a result, recognizing masked faces has become notably complex. Diverse strategies have been introduced to effectively address this obstacle, ensuring consistent recognition performance despite partial facial coverage due to masks or adjuncts like sunglasses.

These methodologies continue to evolve, they exhibit the potential to substantially enhance the accuracy and dependability of face recognition systems in real-world scenarios where facial occlusion represents a prevalent challenge.

Feature extraction serves as a cornerstone in machine learning, accelerating processes and deriving meaningful vectors from original data for downstream use. Within the Masked Face Recognition (MFR) system, optimal feature extraction techniques are pivotal during training and testing. Principal Component Analysis (PCA) diminishes data dimensionality by capturing variance, while transfer learning, exemplified by VGG16 [4] , harnesses pretrained models for domain-specific tasks. Both approaches enhance the MFR system's efficiency and performance by refining data representation and enabling accurate analysis.

PCA (Principal Component Analysis) [5] is a method for streamlining data by extracting essential features from the initial dataset. The primary goal of this technique is to decrease the complexity of multidimensional data while retaining the critical information [6] . Falling under the unsupervised learning category, PCA operates without referencing specific data points. The PCA algorithm executes a linear conversion, reshaping the input data into a novel coordinate system referred to as principal components. These components are linear equations that capture variances and establish a fresh set of variables to represent the data. [6]

Transfer learning is a technique in machine learning that involves utilizing pre-trained models as feature extractors. This approach proves highly flexible by incorporating a pre-trained model for extracting features and performing initial image pre-processing. Notable examples of transfer learning models encompass VGG16, VGG19, ResNet (such as ResNet50, ResNet34, and ResNet18), Inception from the GoogLeNet family, and Xception.

Within the VGG family, models like VGG16 and VGG19 stand out. Similarly, GoogLeNet comprises models like Inception. Additionally, the Residual Network (ResNet) encompasses variations like ResNet50, ResNet34, and ResNet18. These models serve as effective tools for feature extraction and image pre-processing, thus facilitating their integration into divergent models.

DHuang et al. [7] employed ArcFace [8] , a deep network-based facial recognition system, for training using their synthesized dataset. This dataset was created by introducing random occlusions like masks or glasses. In their investigation, the network exhibited the capacity to learn more distinctive features compared to the dataset with only masked faces. Nevertheless, the performance of their approach experienced a significant decline when tested exclusively on the dataset containing masked faces.

Walid Hariri [9] introduced a novel technique centered around eliminating occlusions from facial images using deep learning-based features. Their strategy involved leveraging a pre-trained network to tackle the occlusion-related challenges. By employ a cropping filter technique to excise the occluded portions, specifically those covered by facial masks. This allowed them to retain features solely from the unmasked facial regions. However, it's important to note that the efficacy of the occlusion removal approach is not guaranteed due to the varying positions of facial masks on different individuals. Despite these efforts, the recognition performance of both simulated and real masked facial images by Walid Hariri's method still requires further enhancement.

Recent research has aimed to address Masked Face Recognition (MFR) challenges through the application of attention mechanisms. Li et al. [10] introduced an innovative approach that combines cropping and attention strategies using CBAM (Convolutional Block Attention Module). The cropping-based technique involves eliminating the masked facial area from the input images. Various cropping proportions were investigated to identify the optimal configuration for achieving the highest recognition accuracy. In the attention-based process, reduced weights were assigned to the masked facial features, while higher weights were allocated to the eye vicinity features.

Deng et al. [11] proposed an algorithm named MF-Cosface that tackles MFR using the cosine loss. Their method outperformed the attention-based approach from the previous study. Additionally, they developed an Attention-Inception module, which fused CBAM with Inception-ResNet to enhance focus on the unmasked facial region. Although this technique led to only a marginal improvement in the verification task, it showcased the potential of combining attention mechanisms for refining MFR accuracy. The achieved MFR accuracy reported by Li et al. was approximately 92.61%, while Deng et al. demonstrated enhancements in recognition accuracy through their proposed methods.

Li et al. [12] comprehensive algorithm framework was introduced, comprising two essential modules: de-occlusion and distillation. The de-occlusion module utilizes Generative Adversarial Networks (GANs) to execute masked face completion. This process effectively restores the concealed facial features hidden beneath the mask, mitigating appearance uncertainties. The distillation module, on the other hand, employs a pre-trained model for face classification. In their experiments on the simulated LFW dataset, the highest level of recognition accuracy achieved by Li et al. was approximately 95.44%. This signifies the effectiveness of their algorithm framework in enhancing the accuracy of recognizing masked faces.

Goyal et al. [13] proposed a novel method for recognizing partially masked faces. This method tackles the challenge of identifying individuals even when masks obscure parts of their faces. The innovation lies in employing a bi-orthogonal basis vector approach to create a specialized basis set tailored for masked eigenvectors. The most important point of the approach involves measuring the efficiency of basis representation. This is accomplished by comparing the dimensionality of the space required to accurately represent a facial image using the proposed method versus a conventional orthogonal basis set. The study demonstrates that the suggested technique outperforms the traditional orthogonal basis approach in terms of accuracy, while also achieving a more compact basis representation. By introducing the concept of bi-orthogonal basis vectors and customizing them for masked eigenvectors, the method provides an efficient means of representing facial features. This not only improves accuracy but also showcases its superiority by requiring lower basis space dimensionality compared to traditional orthogonal methods.

Sikha et al [14] proposed a novel approach to masked face recognition through the integration of a cropping-based methodology and an enhanced VGG-16 architecture. This innovative method demonstrates superior performance when compared to prevailing state-of-the-art techniques in the domain of recognizing faces obscured by masks. The authors substantiate the efficacy of their approach through comprehensive experiments on established datasets such as the Georgia Tech Face Dataset, Head Pose Image Dataset, and Face Dataset by Robotics Lab. The proposed methodology comprises four pivotal modules: Face alignment and pre-processing, Mask removal, Feature Extraction, and Face recognition. A significant breakthrough lies in the utilization of a hybrid VGG16-Random Fourier deep learning model, adept at extracting enriched features from the upper facial region while strategically excluding masked portions, thus augmenting subsequent recognition efforts. Moreover, the integration of partial Gappy PCA for eigenvector-based facial reconstruction marks a notable advancement. Through the introduction of optimized image cropping, the proposed model effectively mitigates the computational burden associated with mask detection, offering a comprehensive and promising solution in the realm of accurate masked face recognition.

The author [15] proposed a histogram based recurrent neural network(HRNN) masked face recognition. It uses histograms of oriented gradients(HOG) as the feature extraction process and recurrent neural netwok(RNN) as the deep learning process. The method is designed to solve the problem of undetected masked face images in the recognition system. The experiment is conducted in two datasets Real World Masked Face Dataset(RMFD) and Labeled Face in the Wild simulated Masked Face Dataset (LFW - SMFD). The proposed method achieved a accuracy of 99% and it is designed to overcome the underfitting problems.

E.Alqarella et al. [16] presents an innovative solutions to the challenges posed by facial recognition systems encountering partial and irregular occlusions, such as scarfs or sunglasses. These obstructions can greatly compromise the accuracy of identification processes. To address this issue, the author introduces two novel techniques: "masked projection" and "Masked Correlation Filters (MCFs)." The former likely involves modifying the representation of facial features to incorporate occluded regions, preserving essential information for recognition. The latter technique, MCFs, is likely an advanced method employing correlation filters to match features of obscured faces with templates. The paper demonstrates the effectiveness of these techniques through experimentation on various datasets, revealing remarkable accuracy achievements: the proposed methods yield up to 99.64% accuracy on the Real-World Masked Face (RMFD) dataset, 99.49% on the Simulated Masked Face (SMFD) dataset, and a flawless 100% on the Labeled Faces in the Wild (LFW) dataset. The research suggests that these techniques outperform conventional recognition systems using profile or frontal face images, providing enhanced reliability and accuracy in recognizing masked faces – a significant advancement given the prevalence of face coverings in contexts like the COVID-19 pandemic.

H J Lee et al. [17] tackled the challenge of insufficient masked face images for training face recognition models by introducing a data augmentation technique. They employed this method to create masked face images using a commonly used face recognition dataset. To evaluate the effectiveness of their approach, the author utilized established verification datasets like the masked version of LFW, AgeDB-30, CFP-FP, and a real mask image verification dataset known as MFR2. The experimental results showcased the superiority of their proposed method over the current state-of-the-art technique in masked face recognition (MFR). This suggests that the data augmentation approach significantly enhances the model's ability to recognize faces even when they are partially obscured by masks. In summary, the author addressed the scarcity of masked face images for training face recognition models by introducing a data augmentation method. By applying this technique to a well-known dataset, they demonstrated its effectiveness in generating masked face images. Subsequent evaluations against various verification datasets confirmed that their proposed approach outperforms existing methods, marking a substantial advancement in masked face recognition capabilities.

W Hariri [9] proposed method used the Real-World-Masked-Face-Dataset (RMFRD) to evaluate the performance of the method. The dataset contains 5,000 masked faces and 90,000 non-masked faces. The authors applied an over-sampling technique by cropping some non-masked faces to get an equivalent number of cropped and full faces. The authors used three pre-trained models (VGG-16, AlexNet, and ResNet-50) separately to extract deep features from their last convolutional layers. The quantization technique was then applied to extract the histogram of a number of bins. Finally, Multilayer Perceptron (MLP) was applied to classify faces. The authors used the 10-fold cross-validation strategy to evaluate the recognition performance. The experimental results on the RMFRD dataset showed high recognition performance compared to other state-of-the-art methods.

Chapter 3

METHODOLOGY

The detailed discussion of the methodology followed in the proposed work is presented in this section

3.1 Data Collection and Preparation

In this section, we discuss the data collection process, detailing the two datasets used: the benchmark dataset "COMASK20" [18] and a self-collected dataset.

3.1.1 COMASK20 Benchmark Dataset

The "COMASK20" dataset is a widely used benchmark in masked face detection. It contains a diverse collection of facial images, each categorized as either "masked" or "not masked." This dataset serves as a foundational resource for training and evaluating models for masked face detection. It helps to establish a baseline for performance and enables comparisons with other methods. Some samples from the COMASK20 dataset is shown in 3.1

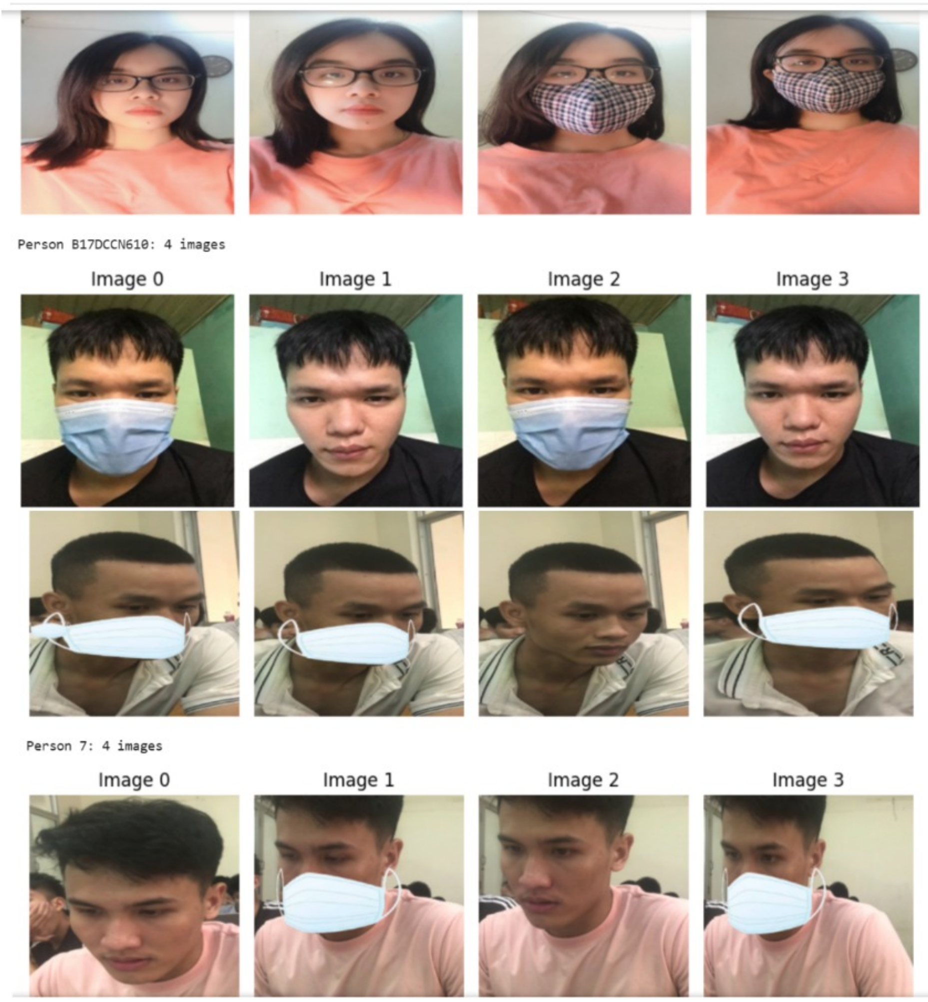


Fig. 3.1 SAMPLE IMAGES FROM COMASK20 [18] DATASET

3.1.2 Self-Collected Dataset

To enhance the model's performance and adaptability to various scenarios, we collected a self-made dataset. This dataset includes facial images of 50 individuals, with each individual forming a distinct class. For each class, we captured images under varying lighting conditions, poses, and orientations, encompassing both masked and unmasked instances. This diverse dataset aims to improve the model's ability to handle real-world scenarios and different face variations.

3.1.3 Data Augmentation

Data augmentation is a crucial step in enhancing the capabilities of our masked face recognition model. By artificially expanding the dataset through various transformations, we aim to expose the model to a wider range of scenarios it might encounter in real-world situations.

Augmentation Techniques

We applied a variety of augmentation techniques to both the "COMASK20" benchmark dataset and the self-collected dataset:

- **Rotation**: Images were randomly rotated by a certain degree. This simulates variations in head poses and helps the model learn to recognize faces from different angles.
- **Horizontal Flipping**: Images were horizontally flipped, effectively doubling the dataset's size while introducing mirror-symmetrical variations. This helps the model learn features that are invariant to left-right orientation.
- **Brightness and Contrast Adjustments**: Random adjustments to brightness and contrast were applied. This accounts for varying lighting conditions that the model might encounter in real-world scenarios.

- **Zoom**: Random zoom-in and zoom-out transformations were used to simulate varying distances between the camera and the subject. This helps the model learn to recognize faces at different scales.

Rationale and Parameters

Each augmentation technique was applied with carefully chosen parameters. For instance, rotation angles were limited to realistic head poses, and the magnitude of brightness adjustments was controlled to prevent over-saturation or under-exposure. The goal was to strike a balance between introducing useful variations and ensuring the generated samples remained plausible and relevant.

Impact on Model Training

By employing these augmentation techniques, we expose our model to a wider range of possible inputs. This helps prevent overfitting, as the model learns to generalize from the diverse augmented dataset rather than memorizing specific instances. Additionally, the model becomes more adept at recognizing masked faces under different conditions, contributing to its overall accuracy and reliability.

In conclusion, data augmentation is a pivotal component of our masked face recognition pipeline. It empowers our model with the ability to handle various real-world scenarios and challenges by creating a more comprehensive and diverse training dataset.

Data Balancing

In both datasets, care was taken to ensure balanced representation of classes. Imbalanced datasets can lead to biased models that perform well on the majority class but poorly on the minority class. To counter this, images were distributed proportionally across classes, considering both masked and unmasked instances.

3.1.4 Dataset Splitting

To evaluate the model's performance, the datasets were split into training, validation, and testing sets. The training set was used to train the model, the validation set for hyperparameter tuning, and the testing set to assess the model's generalization on unseen data.

In summary, the combination of "COMASK20" and our self-collected dataset, along with careful data augmentation and balanced class representation, forms a comprehensive dataset for training and evaluating our masked face detection model.

3.2 Proposed Methodology

The proposed workflow starts with the input dataset containing various face images. The images undergo face extraction using the Python Imaging Library (PIL), ensuring accurate isolation of the facial regions. Once the faces are extracted, FaceNet, a deep learning model, is employed to generate high-dimensional face embeddings that capture distinct facial features. These embeddings are then input into a Support Vector Classifier (SVC), a specialized machine learning algorithm, which classifies the embeddings into categories such as masked and unmasked faces. The output of the SVC provides the recognition of masked faces and corresponding results 3.2

3.2.1 Face detection

Face detection is a critical step in computer vision and facial recognition systems. It involves identifying and localizing the presence of human faces within an image or a video frame. The primary objective of face detection is to determine whether there are any faces present in the given visual data and to provide the coordinates of the bounding boxes that enclose each detected face. This step is crucial in all the face recognition, emotion analysis and human-computer interaction.

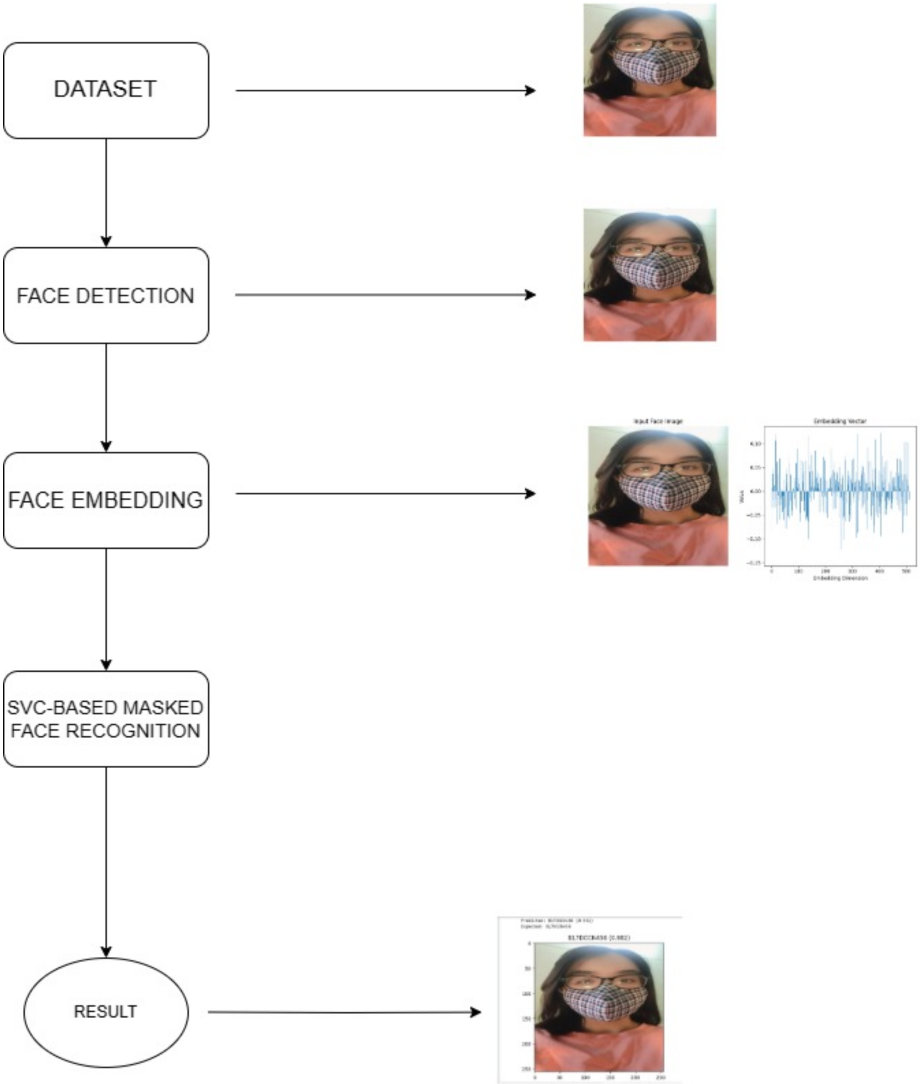


Fig. 3.2 Pictorial representation of proposed methodology for masked face recognition

3.2.2 Face Embeddings

Face embeddings involve converting facial images into numerical vectors, typically with dimensions such as 128-d or 512-d. This process captures the unique facial features in a structured numerical format. Notably, the utilization of FaceNet [3] , a pioneering face recognition model, brings several advantages to this domain.

One advantage of utilizing FaceNet's face embeddings is their ability to represent intricate facial characteristics with remarkable precision. By mapping facial images into high-dimensional vectors, subtle details crucial for accurate recognition are encapsulated, enhancing the model's capacity to distinguish between individuals.

Moreover, FaceNet's embeddings enable efficient and rapid comparisons between faces. Through numerical representations, similarity measures can be quickly computed, facilitating tasks like face verification and identification across large datasets. This expeditious processing is particularly valuable in scenarios where real-time or near-real-time performance is critical.

Furthermore, FaceNet's embeddings promote robustness against variations in lighting, pose, and other environmental factors. The embedding vectors are adept at capturing essential identity-related information while being less sensitive to external influences. This resilience enhances the reliability of face recognition systems across diverse settings and conditions.

The adoption of FaceNet's face embeddings empowers the conversion of facial images into compact, discriminative, and versatile numerical vectors. These embeddings excel in accurately representing facial features, facilitating rapid comparisons, and maintaining performance consistency amidst environmental variations.

The FaceNet architecture involves training a deep neural network to learn discriminative embeddings using loss functions like the Triplet Loss. The Triplet Loss encourages the network to pull embeddings of the same person's face closer together while pushing embeddings of different people's faces farther apart in the embedding space.

The Triplet Loss is given by 3.1

$$L_{\text{triplet}}(A,P,N) = \max\left(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0\right) \tag{3.1}$$

Where: - $f(\cdot)$ represents the embedding function. - A is the anchor sample. - P is the positive sample. - N is the negative sample. - α is the margin that helps define the difference required between positive and negative pairs.

The embeddings are optimized during training to minimize this Triplet Loss, resulting in embeddings that can accurately capture facial features and distinguish between individuals.

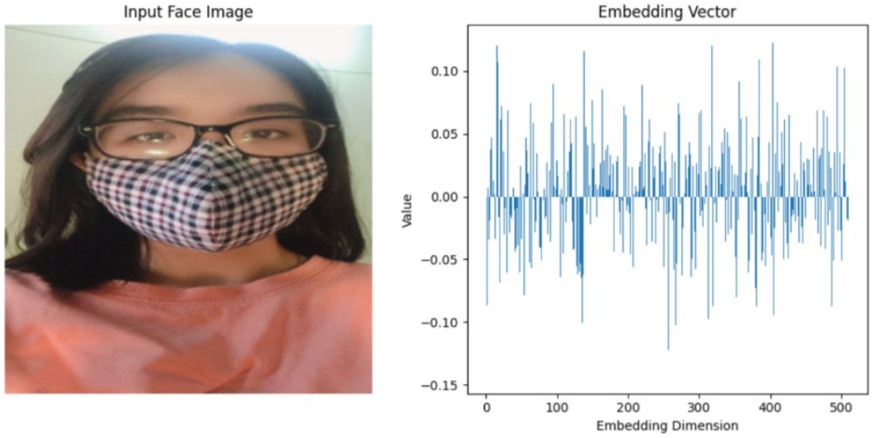
The FaceNet model takes an input facial image and generates an embedding vector using the trained neural network. This embedding vector is designed to capture essential features that define an individual's face.

During training, the model learns to make the embeddings of the same person's face closer together in the embedding space and push embeddings of different people's faces apart. This process helps the model to create a meaningful embedding space where faces of the same person are clustered together.

When performing face recognition tasks, new facial images can be converted into embeddings using the same learned network. By measuring the distance between embedding vectors, one can determine the similarity between faces. If the distance between two embeddings is small, the faces are likely from the same person. Conversely, if the distance is large, the faces are likely from different people.

This process enables tasks like face verification (determining if two images are of the same person) and face identification (matching a face against a database of known individuals) using the similarity between embedding vectors.

The sample output of the FaceNet embedding is in figure 3.3



A) B)

Fig. 3.3 A) input image captures a person's face, possibly with different expressions and masked or unmasked. B) After FaceNet, an intricate embedding vector emerges, condensing unique facial features into numerical values, crucial for streamlined face recognition.

3.2.3 Support Vector Classifier in Masked Face Recognition

The Support Vector Classifier (SVC) [2] is a powerful machine learning algorithm commonly used for classification tasks, even when dealing with complex, non-linear data. In the context of face detection with masks, the SVC is harnessed to determine whether a person's face is detected with a mask or without.

The SVC operates by finding the optimal decision boundary that separates different classes of data. In cases of non-linear data, a linear boundary might not suffice. This is where kernels come into play. Kernels are functions that map data into higher-dimensional spaces, often revealing linear separability that wasn't apparent in the original feature space. The SVC exploits these kernels to find a linear decision boundary in the transformed space, effectively addressing non-linearity.

Mathematically, given a set of data points $\mathbf{x}_i$ and corresponding labels y_i, the decision boundary can be represented as $\mathbf{w} \cdot \mathbf{x} + b = 0$, where $\mathbf{w}$ is the weight vector and b is the bias term. In the case of non-linear kernels, the decision boundary can be expressed as:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b,$$

where α_i are the learned coefficients, y_i are the labels, N is the number of support vectors, and $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function that computes the similarity between data points $\mathbf{x}_i$ and $\mathbf{x}$ in the transformed space.

Kernels play a pivotal role in enhancing the Support Vector Classifier's (SVC) ability to address the intricate challenges posed by masked face recognition. Masked face recognition involves identifying faces even when obscured by masks, demanding the identification of nuanced patterns that might not be discernible in conventional data representations. Kernels offer a compelling solution to these challenges.

The inherent power of kernels lies in their capacity to capture intricate relationships that are prevalent within complex data. For tasks like masked face detection, where the presence or absence of masks can lead to intricate and non-linear variations in features, kernels enable the SVC to create decision boundaries that flexibly adapt to these complexities. This adaptability proves to be a critical asset in tackling scenarios that demand a deep understanding of intricate relationships within the data.

The kernel trick, a fundamental concept, facilitates the transformation of input data into higher-dimensional spaces. By doing so, the SVC can unveil relationships that remain concealed in the original feature space. In the realm of masked face recognition, where the interplay between facial features and masks might not be immediately evident, this transformation allows the SVC to discern nuanced patterns that are crucial for accurate identification.

A distinguishing feature of the SVC equipped with kernels is its ability to effectively handle non-linear data. Unlike linear classifiers, the SVC's utilization of kernels empowers it to navigate complex relationships, positioning it as an invaluable tool for tasks like masked face recognition that demand the capture of intricate and non-linear connections within the data.

In tandem with addressing non-linearity, kernels contribute significantly to enhancing the generalization prowess of the model. The SVC's capability to grasp intricate patterns through kernels translates to a superior ability to avoid overfitting. As the SVC learns more meaningful features rather than simply memorizing training data, it can generalize well to new and unseen examples, a critical aspect in achieving reliable masked face recognition results.

However, the choice of the kernel function remains paramount in the context of masked face recognition. Different kernels are tailored to different data characteristics and relationships. Selecting the most appropriate kernel function becomes an art, crucial for unlocking the full potential of the SVC in capturing the complexity of masked face patterns.

In essence, the synergy between kernels and the SVC is a cornerstone of success in complex classification tasks like masked face recognition. Kernels equip the model to unravel non-linear patterns embedded within the data, thereby elevating its accuracy and adaptability while averting overfitting risks. Through this mechanism, the SVC emerges as a potent solution for the intricate challenges posed by masked face recognition, demonstrating the pivotal role that kernels play in pushing the boundaries of modern machine learning.

Kernel Selection and Training Set

In the context of masked face recognition using the Support Vector Classifier (SVC), the selection of an appropriate kernel significantly impacts the model's performance. One frequently employed kernel is the polynomial kernel. The polynomial kernel computes the similarity between pairs of data points using polynomial functions, enabling the SVC to capture intricate relationships beyond linear separability.

When applying the polynomial kernel to masked face recognition, the resultant decision boundary is shaped by polynomial functions based on the features of the input data. This is especially advantageous for addressing the challenges posed by masked face recognition. The presence of a mask introduces complex feature patterns that may not be easily separable using linear methods. The polynomial kernel's capability to capture higher-order interactions makes it a suitable choice for effectively addressing this task.

In the training phase of masked face recognition, a diverse set of images is typically employed. This dataset comprises images of individuals, each considered as a single class, with some individuals wearing masks and others without masks. This diversity ensures that the model learns a wide array of patterns and features associated with both masked and unmasked faces. Maintaining a balanced representation of individuals with masks and without masks within the training data is important. This balance is crucial for enabling the model to generalize effectively to new and unseen data, ensuring robust performance in real-world scenarios.

Chapter 4

RESULTS AND DISCUSSIONS

4.1 Evaluation Methods (cross validation)

In the context of evaluating machine learning models, [19] cross-validation is a vital technique. This approach involves dividing the dataset into multiple folds, enabling iterative training and evaluation of the model. The primary purpose of cross-validation is to estimate the model's performance on unseen data.

Implementation of Cross-Validation

In this study, a 5-fold cross-validation strategy was adopted. The dataset was divided into five approximately equal-sized subsets, maintaining the original class distribution. The model underwent five training iterations, using four folds for training and one for validation in each iteration. Performance metrics such as accuracy, precision, recall, etc., were computed for each iteration based on predictions made on the validation set. Final performance metrics were derived by averaging the metrics obtained from all five iterations.

The results are summarized in Table 4.1 . We observed high accuracy levels for both datasets, with COMASK20 achieving an accuracy of 97.2% and SelfCollected achieving an accuracy of 89.4%.

Table 4.1 5-Fold Cross-Validation Results

Dataset	**Accuracy (%)**	**F1 Score**
COMASK20	97.2	97.5
SelfCollected	89.4	88.9

4.2 Evaluation Metrics

In addition to accuracy, other important evaluation metrics [20] were calculated to provide a comprehensive assessment of the model's performance:

- **Precision:** Precision measures the proportion of correctly predicted positive observations out of the total predicted positives. In other words, it indicates how many of the instances predicted as positive by the model are actually positive, helping to quantify the model's ability to minimize false positives. A high precision value suggests that the model is careful in its positive predictions, reducing the chances of labeling negatives as positives. The calculation is done by 4.1

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{4.1}$$

- **Recall (Sensitivity or True Positive Rate):** Recall calculates the ratio of correctly predicted positive observations to the total actual positives in the dataset. In essence, it highlights the model's ability to identify all positive instances, regardless of how many false negatives might be produced. High recall suggests that the model is adept at capturing positive instances, minimizing the chance of overlooking them. It is calculated by 4.2

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{4.2}$$

- **F1-Score:** The F1-score is the harmonic mean of precision and recall. It provides a balanced measure that combines the strengths of both metrics. F1-score is particularly useful when the class distribution is imbalanced, as it gives equal importance to precision and recall, helping to assess the model's overall performance across the positive class. A higher F1-score indicates a better balance between precision and recall. F1score is calculated using 4.3

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{4.3}$$

4.3 Comparison

To contextualize the performance of the developed masked face recognition system, a comparison was made with existing methods on the COMASK20 dataset with and without cross validation. The tabulation of both evaluation metrics were shown in 4.2 and 4.3 .The results indicate that our proposed approach achieved competitive performance.

Table 4.2 Comparison of the proposed method with existing method before cross-validation

Authors	Dataset	Recall	Precision	F1-Score
Proposed	COMASK20	99.8	99.8	99.6
Proposed	Self-Collected	97.4	98.3	97.2
Vu et al.	COMASK20	87.0	87.0	87.0

Table 4.3 Comparison of the proposed method with existing method after cross validation

Authors	Dataset	Recall	Precision	F1-Score
Proposed	COMASK20	97.6	98.5	97.5
Proposed	Self-Collected	89.4	92.8	88.9
Vu et al.	COMASK20	87.0	87.0	87.0

4.4 Ethical Considerations

In the deployment of any facial recognition system, ethical considerations are crucial. Potential biases, privacy concerns, and unintended consequences must be thoroughly examined. While this project focuses on masked face recognition, it is essential to acknowledge and address these ethical concerns, ensuring that the technology is used responsibly and with due respect for individual rights. The COMASK20 dataset is taken from [18] and the self collected dataset is not provided any where just used for research purpose only.

4.5 Test Images

The model is created as the proposed methodology and some random test samples were tested like expected and predicted and the model predicted the persons correctly and there are some wrong predictions also, the resultant images is shown in figure 4.1

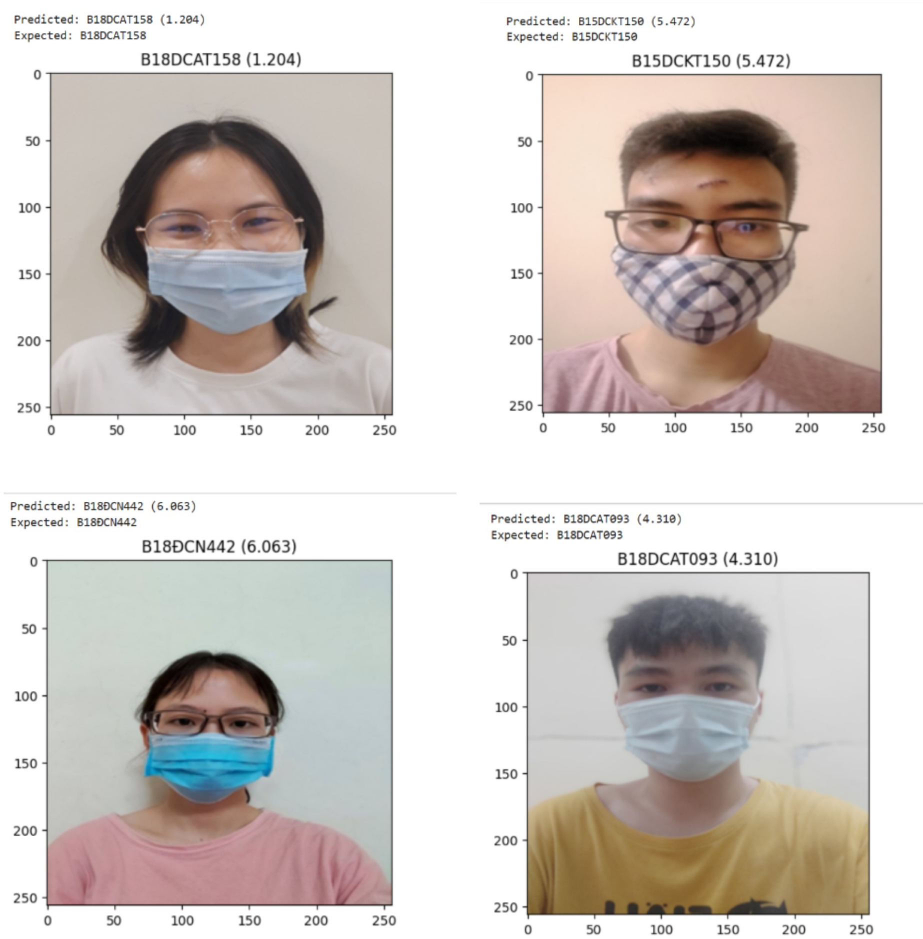


Fig. 4.1 SAMPLE PREDICTION

Chapter 5

CONCLUSION

In conclusion, the application of masked face recognition using Support Vector Machines (SVM) with FaceNet embeddings has demonstrated promising results on two distinct datasets: the benchmark COMASK20 dataset and a self-collected dataset. The achieved accuracies of 97.2% and 89.4% on these datasets, accompanied by F1 scores of 97.5% and 88.9% respectively, underscore the effectiveness of the proposed approach in handling the challenging scenario of masked face recognition. The experiment's success highlights the potential of combining advanced facial recognition techniques like FaceNet with machine learning classifiers like SVM, even in scenarios where face masks significantly obscure facial features. The high accuracy and F1 scores are indicative of the model's robustness in distinguishing individuals despite the presence of masks, which is crucial in real-world applications such as security, access control, and public health management.

However, there are several avenues for future research that could enhance the current findings. Firstly, exploring the utilization of other feature extraction methods apart from FaceNet, such as deep learning-based architectures or even traditional methods, could provide valuable insights into the comparative performance and the potential trade-offs involved. Additionally, investigating the impact of varying mask types, colors, and coverage on the model's performance would contribute to a more comprehensive understanding of its generalization capabilities. Furthermore, the experiment's scope could be extended to include larger and more diverse datasets to assess the model's adaptability to a broader range of scenarios and populations. This might involve exploring cross-dataset generalization and potential biases that could affect real-world deployment. Incorporating real-time aspects into the research, such as handling dynamic scenarios where individuals are donning or removing masks, could provide valuable insights into the model's practical applicability. Additionally, investigating ensemble methods, hybrid architectures, or even semi-supervised learning approaches might lead to further performance improvements.

In conclusion, the success of the masked face recognition using SVM with FaceNet embeddings is a promising step towards addressing real-world challenges posed by face masks. However, continuous exploration of new techniques, evaluation on diverse datasets, and adaptation to evolving scenarios will be key for refining and extending the model's capabilities for practical deployment.

BIBLIOGRAPHY

References

[1] H. N. Vu, M. H. Nguyen, and C. Pham, "Masked face recognition with convolutional neural networks and local binary patterns," *Applied Intelligence*, vol. 52, no. 5, pp. 5497–5512, 2022.

[2] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[4] A. B. Perdana and A. Prahara, "Face recognition using light-convolutional neural networks based on modified vgg16 model," in *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)*. IEEE, 2019, pp. 1–4.

[5] S. Roweis, "Em algorithms for pca and spca," *Advances in neural information processing systems*, vol. 10, 1997.

[6] M. Verleysen and M. Verleysen, "Principal component analysis (pca) principal component analysis (pca) principal component analysis (pca)," *Stat. June*, vol. 2007, pp. 1–8, 2001.

[7] B. Huang, Z. Wang, G. Wang, K. Jiang, K. Zeng, Z. Han, X. Tian, and Y. Yang, "When face recognition meets occlusion: A new benchmark," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4240–4244.

[8] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4690–4699.

[9] W. Hariri, "Efficient masked face recognition method during the covid-19 pandemic," *Signal, image and video processing*, vol. 16, no. 3, pp. 605–612, 2022.

[10] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Self-restrained triplet loss for accurate masked face recognition," *Pattern Recognition*, vol. 124, p. 108473, 2022.

[11] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, "Mfcosface: a masked-face recognition algorithm based on large margin cosine loss," *Applied sciences*, vol. 11, no. 16, p. 7310, 2021.

[12] Y. Li, S. Liu, J. Yang, and M.-H. Yang, "Generative face completion," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3911–3919.

[13] N. Goyal, D. Mahor, and K. Khare, "Recognition of masked faces," *Journal of Optics*, pp. 1–9, 2022.

[14] O. Sikha and B. Bharath, "Vgg16-random fourier hybrid model for masked face recognition," *Soft Computing*, vol. 26, no. 22, pp. 12 795–12 810, 2022.

[15] W.-J. L. Chong, S.-C. Chong, and T.-S. Ong, "Masked face recognition using histogram-based recurrent neural network," *Journal of Imaging*, vol. 9, no. 2, p. 38, 2023.

[16] E. Alqaralleh, A. Afaneh, and Ö. Toygar, "Masked face recognition using frontal and profile faces with multiple fusion levels," *Signal, Image and Video Processing*, vol. 17, no. 4, pp. 1375–1382, 2023.

[17] V. Pann and H. J. Lee, "Effective attention-based mechanism for masked face recognition," *Applied Sciences*, vol. 12, no. 11, p. 5590, 2022.

[18] H. Vu, M. Nguyen, and C. Pham, "Masked face recognition with convolutional neural networks and local binary patterns," *Applied Intelligence*, 08 2021.

[19] S. Bates, T. Hastie, and R. Tibshirani, "Cross-validation: what does it estimate and how well does it do it?" *Journal of the American Statistical Association*, pp. 1–12, 2023.

[20] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.