Journal of Computer Science and Engineering in Innovations and Research
RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

Machine-learning-Driven Masked Face Recognition: Boosting Accuracy with Refined Facenet Model

Anusua Basu[1*], Anjan Choudhury[2]

[1] Department of Computer Science Engineering, University of Kalyani, West Bengal, India and mail2anusuacse@gmail.com
[2] Department of Robotics and Additive Manufacturing Technology, Maulana Abul Kalam Azad University of Technology, West Bengal, India

Abstract— Face masks significantly impair traditional face recognition systems by obscuring crucial facial features like the nose, mouth, and chin. The variety of mask types, including surgical and cloth masks, further complicates recognition due to differences in materials, shapes, and colors. This research aims to address the challenges posed by facial occlusion, particularly mask-wearing, and improve recognition system performance. The study leverages the Facenet deep learning model, which uses a convolutional neural network (CNN) to generate facial embeddings for identification or verification purposes. To overcome the limitations of existing models in recognizing masked faces, a Refined Facenet model is proposed. The experiment compares three models Facenet, Refined Facenet, and VGG16 paired with five classifiers: Support Vector Machine (SVM), Decision Tree, k-Nearest Neighbors (KNN), Random Forest, and Gaussian Naïve Bayes. Datasets used include Unmasked, Masked, Merged, and Augmented versions of these datasets, with evaluations performed using K-fold cross-validation and Bootstrap sampling. Results demonstrate that the Refined Facenet model with KNN consistently achieves higher accuracy and optimized CPU execution time across different datasets, providing a more robust solution for face recognition under occlusion conditions.

Keywords*— Image Processing, Convolutional Neural Networks (CNNs), kNearest Neighbors (KNN), Machine Learning, Facenet, Face *Recognition

I. INTRODUCTION

After the COVID-19 coronavirus epidemic, facial masks have been worn by almost everyone, which has posed a huge challenge to deep face recognition. However, the following problems are caused by wearing a mask: Masks are taken advantage of by fraudsters and thieves, with crimes being committed and identities not being recognized [1]. Community access control and face authentication have been made very difficult tasks when a large part of the face is hidden by a mask [2]. Existing face recognition methods have not been efficient when masks are worn, as a complete face image is not provided for description [3]. The exposure of the nose region is deemed very important in face recognition, as it is used for face normalization, pose correction, and face matching [4].

For biometric authentication, face recognition is widely utilized due to its convenience and feasibility. However, the importance of addressing challenges posed by face masks has been underscored by the COVID-19 [35] pandemic, as recognition systems can be hindered. Performance is also impacted by factors like pose variations, illumination changes, aging, and facial expressions, with occlusion presenting a significant hurdle in practical scenarios due to environmental conditions, camera angles, and subject movement. To ensure robust facial recognition across diverse conditions, enhancements to existing methods are required [36]. The development of adaptable systems capable of maintaining accuracy and privacy amidst evolving global health and societal norms is sought, guiding future research and innovation in this field.

In computer vision, face recognition is regarded as a critical problem with numerous real-world applications. It typically involves four steps: face detection, facial landmark detection, facial feature extraction, and classification. Significant improvements in object detection have been made by deep learning, leading to the success of models like MTCNN. Various approaches for masked face recognition have been introduced in recent studies. Mandal and Okeukwu fine-tuned a pre-trained ResNet-50 model, with a focus on recognizing masked faces through domain adaptation and occlusion handling. A combination of ResNet-50 with YOLO v2 was proposed by [33] achieving 81% precision for mask detection, while an enhancement of tiny YOLO v4 [34] with spatial pyramid pooling was made by reaching 84% precision. Despite its accuracy, VGG16 [5] suffers from longer execution times and reduced accuracy for occluded faces. Suboptimal accuracies (80%-85%) have been

exhibited by current methodologies, along with a lack of comparative analysis, with potential improvements expected to lead to higher CPU utilization.

Challenges lie in the datasets used for training models, which contain images of individuals in various conditions like low light, tilted angles, and masked or covered faces. Firstly, the dataset was trained using the pre-trained facenet model, but unsatisfactory accuracies were achieved, and much time was consumed in training. Secondly, changes were made to the existing facenet model by adjusting the activation function and momentum values, as well as replacing the triplet loss function with proxy loss. Through these enhancements, CPU time was significantly reduced, and accuracies were increased.

II. RELATED WORK

Deep learning models have been wonderful, robust, and correct in Face Recognition for a couple of years. In this section, a literature review of work done during the years 2014 till 2024 on Convolutional Neural Network (CNN) Model for Face Recognition is shown in the Table 1.

Table 1: Review on CNN methods for Face Recognition from the year 2014 to 2024.

Years	Methods	Applications
2014	VGG (VGG-11, VGG-13, VGG-16, VGG-19): Convolutional Neural Networks [5].	General-purpose feature extraction in face recognition, with variations for depth, accuracy, and computational efficiency
	VGG-11 [5]	Real-time face recognition with modest depth for quicker training and inference
	VGG-13 [5]	Feature extraction with a balance between accuracy and computing efficiency
	VGG-16 [5]	Widely used for face recognition, resilient in varying lighting and position scenarios
	VGG-16 FaceNet Adaptations [5]	High recall and precision tasks, security, and authentication systems
	VGG-19 [6]	Improved feature extraction, commonly used in transfer learning, especially for fine-tuning on face recognition after pre-training on ImageNet
2015	FaceNet: Deep Neural Network + Triplet Loss [7].	Face recognition and clustering
2017	SphereFace: Angular Softmax Loss, Hypersphere Embedding [8].	Face recognition
2018	CosFace: Cosine Margin Loss, Deep Residual Networks [9].	Deep face recognition

2019	ArcFace: Additive Angular Margin Loss, Residual Networks [10].	Masked and unmasked face recognition
	Deng and Feng: MFCosface, Large Margin Cosine Loss [11].	Masked-face recognition algorithm, face image restoration.
2020	AdaFace: Adaptive Margin Loss [12].	Robust face recognition under varying conditions, including occlusions and expressions
2021	CurricularFace: Dynamic Curriculum Learning [13].	Face recognition with emphasis on hard examples and large intra-class variations
2021	MagFace: Magnitude-Aware Feature Embedding [14].	Face recognition focusing on quality-aware embeddings, suitable for large-scale recognition tasks
2022	PartialFace: Partially Occluded Face Recognition [15].	Recognition of faces with partial occlusions (e.g., masks, sunglasses)
2023	MaskedFaceNet: Dual-Stream Network for Masked Face Recognition [16].	Refined recognition for masked faces using a dual-stream network architecture
	PoseFace: Pose-Invariant Face Recognition [17].	Robust face recognition across various pose angles
2024	ResNet-FR: Residual Networks for Face Recognition [18].	State-of-the-art face recognition using deep residual networks, optimized for large-scale datasets
	Dual-Teacher: Knowledge Distillation for Face Recognition [19].	Improved face recognition using dual-teacher knowledge distillation methods

This is a research gap that a lot of contribution has been made according to the literature reviews to address this challenge. The restoration-based techniques were proposed as novel methods in a face recognition. However, different conditions like illumination, occluded objects, and segmentation output of detected masks made the restoration method sensitive. As a result, an imprecise face image was produced, resulting in recognition accuracy decline. Afterward, transfer learning was used to fine-tune the pretrained networks using different strategies and datasets. The exploration to find the optimal model configuration suited for masked or occluded face recognition.

III. NOVELTY OF THE PROPOSED WORK

The novelty of the proposed work is found face recognition arises from the need to address challenges associated with partially obscured facial features, such as those caused by masks or other obstructions. To significantly enhance the accuracy of facial recognition systems when occlusions are present, while simultaneously optimizing the detection time. This work is particularly relevant today, as mask-wearing for health reasons has become commonplace, affecting the performance of conventional face recognition technologies. Originality of the work is to address seek to improve facial recognition performance even when faces are partially covered, like with masks, by reducing resource consumption and minimizing detection time. Additionally, this system can play a vital role in security applications, where accurate face recognition remains critical despite limited visibility of facial features. Recognizing that deep learning models require extensive datasets to perform effectively, also addressed the limitations of our dataset, which contains only a few hundred images, compared to millions used by other systems. To expand our dataset, utilized Generative Adversarial Networks (GANs) to generate synthetic images a process that will be discussed further in subsequent sections.

A. Material and methods

In this research, the Masked Face Recognition Dataset was used, including both real and augmented data, generated through GANs, to enhance training diversity. The proposed

method employs a refined Facenet model, leveraging the InceptionResNetV2 architecture and Proxy Anchor Loss to improve recognition accuracy, particularly for masked faces, and optimize resource consumption and execution time.

B. Dataset

The public dataset [32] used for the data modelling process is the Masked Face Recognition Dataset, which includes variations in pose, illumination, resolution, and different types of masks worn by subjects. Initially, three datasets were used, and three more datasets of augmented images were added. Unmasked Dataset has 658 images of 15 individuals with around 30 images of each individual. The Masked Dataset is composed of 672 images with approximately 40 images per person. The Merged Dataset is an aggregated data set of 1,330 images, consisting of masked and unmasked images of each person (approximately 80 images per person). Generative Adversarial Networks (GANs) were used to create augmented datasets. The Augmented Unmasked Dataset includes 1,410 images and 90 about images per person, and the Augmented Masked Dataset includes 1095 images and about 70 images per person. The Augmented Merged Dataset consists of72000images, for about 140 per person. For each dataset, there are 15 classes (15 faces), arrangement of 3 to 20 images per class. The combined dataset contains both unmasked images and masked images. These accuracies were evaluated and analyzed by performing training and testing on each dataset and systematically comparing the resultant accuracies.

IV. METHODOLOGY

In this part, we briefly discuss how to enhance facenet model. Then it will explain how to append proxy anchor to boost the performance. Next, we describe the key components of the proposed network design.

A. Refined Facenet Model

The Refined facenet model uses the deep neural network InceptionResnetV2 [21] to provide embeddings for the face images. The InceptionResNetV2 model comprises 21 blocks, each integrating convolutional layers, batch normalization, activation functions, and residual connections. A Conv2D layer with 32 filters, a 1x1 kernel size, and "same" padding reduces input dimensionality. Batch normalization, with a momentum of 0.99, stabilizes and accelerates training. The 'leaky_relu' activation introduces non-linearity. The output from three branches is concatenated along the depth axis (axis=3), and a 1x1 convolution combines the feature maps into a single representation. A Lambda layer scales the output, followed by an Add operation, where the scaled result is added back to the original input, maintaining input data and helping address vanishing gradient issues. The final activation applies non-linearity to the combined output, enhancing the network's learning ability.

The model also incorporates two reduction blocks for down sampling, efficiently combining convolutions, pooling, and concatenation to capture distinct input properties. A classification block includes global average pooling, which summarizes each feature map into a single value, and a dense layer with 128 units for feature compression. Dropout (20%) prevents overfitting by randomly deactivating neurons during training, and batch normalization stabilizes activations. This architecture allows the network to efficiently learn complex patterns while maintaining stability and generalization capabilities on new data. Each face image is mapped into a Euclidean space such that the distances in that space corresponds to similarity in face, the similarity between two face images is assessed based on the squared L2 distances [20]. Now, the Proxy Anchor loss function have been used to calculate the distance between embeddings and train the neural network learnable parameters. The Refined Facenet model structure incorporates optimized deep convolutional layers and innovative loss functions to improve facial recognition accuracy and robustness.

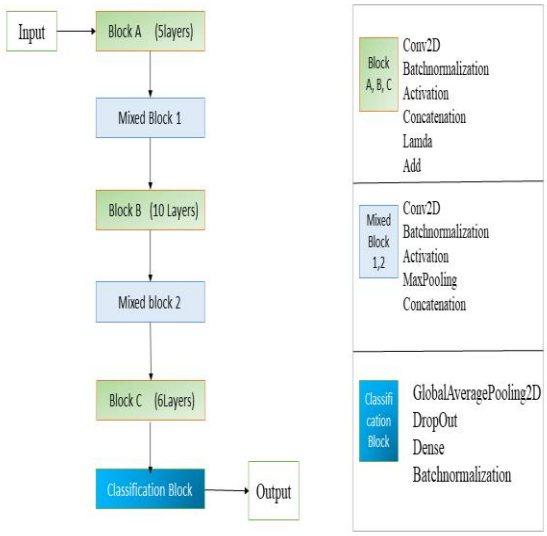


Figure 1: Refined Facenet Model

InceptionResNetV2 [21] is a deep and efficient model that integrates both the benefits of Inception network architectures and universal residual connections to yield superior accuracy with a smaller number of parameters. It receives images of Shape 160 x 160 x 3 as input and creates 128-dimensional image embeddings. These embeddings were then L2 normalized so they are on the same scale for loss calculation. In this model we use the Proxy Anchor Loss, meaning that the loss is calculated based on the embeddings produced by the deep neural network and that the proxies of each one of the classes are updated as well as the network learnable parameters. The model was trained using the ADAM [22] (Adaptive Moment Estimation) optimizer, which is an advanced stochastic gradient descent method that updates network weights iteratively based on training data and Proxy Anchor Loss [20]. The use of 3×3 Inception modules combined with residual connections in InceptionResNetV2 enables the model to be computationally efficient while also achieving state-of-the-art results.

B. Flowchart

This is Figure 2: Image Classification Using a Refined Facenet Model with Proxy Anchor Loss It summarizes the important stages starting from the processing of the input image until its classification.

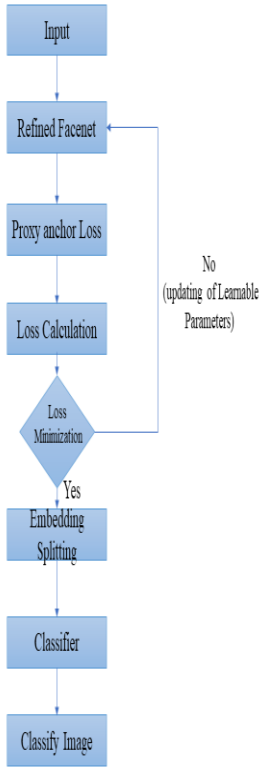


Figure 2: Workflow of Face Recognition using Refined Facenet Model.

The figure 2 illustrates the process of image classification using a refined Facenet model and proxy anchor loss. The process begins with an input image that is passed through the refined Facenet, where feature extraction takes place. The proxy anchor loss function is then applied to measure embedding distances, improving the model's feature learning. The calculated loss is minimized through backpropagation, refining the model's parameters. After loss minimization, the embeddings are split, and these are sent to a classifier. Finally, the classifier assigns the input image to a specific category based on the learned features.

C. Loss Function and Parameters

Proxy Anchor loss [20], different momentum value and different activation function used in the proposed model that are described in below.

Networks are trained to project data onto an embedding space where semantically similar data are closely grouped. Proxy-based loss introduces representative points, or proxies, for each class, simplifying training by using these proxies instead of computing pairwise distances between all samples. The loss pulls data of the same class close to their proxy and pushes others away, facilitating interaction between data through the proxy anchors during training.

Given a dataset with N samples and C classes, the Proxy Anchor loss can be formally defined as follows:

- **Proxies:** For each class $c \in \{1,2,\ldots,C\}$ $c \in \{1,2,\ldots, C\}$, a proxy $PcPc$ is defined. These proxies are learned along with the model parameters.
- **Embeddings:** Let x_i be the embedding of the i-th sample and yi be its corresponding class label (encoded in the one hot label encoder).

As compared to Triplet Loss, Proxy Anchor avoids the need for computing distances between all pairs of samples by using proxies, which reduces computational complexity from (N^3) to $(N \cdot C)$, where N is the number of samples in a batch and C is the number of classes.

The momentum hyperparameter [23] is a crucial role in optimizing Convolutional Neural Networks (CNNs) using gradient descent. It smooths out parameter updates to make them more stable, and to smooth out historical gradients alongside the current gradient, momentum facilitates more consistent convergence to an optimal solution. It also prevents the optimization process from getting stuck in local minima or fluctuating excessively, particularly in areas of flat or gently sloping loss surfaces. While a carefully chosen momentum value, such as 0.99, can expedite convergence, avoid overfitting by not settling prematurely into local minima of the training data, and therefore improve the model's generalization capability. But very high momentum values can lead to instability during training, and may cause overshooting of optimal parameter values.

For this project work, five machine learning classifiers, namely, Support Vector Machine [24] (SVM), Decision Tree [25], K-Nearest Neighbors [26] (KNN), Random Forest[27], and Naive Bayes [28], are used. They are used for different aims, including high-dimensional spaces, recursive data partitioning, instance-based learning, ensembles of different decision trees and classification through Bayes' Theorum. Different classifiers come with different advantages in terms of the ability to learn non-linear relationships in the data, preventing overfitting, and resulting in more accurate predictions.

Train/valid split: A very common technique to assess the performances of a model and its generalization is k-fold crossvalidation [29], where a dataset is split k-folds, training on k-1 folds and validating on the left-out fold in a repeated manner. This approach minimizes dataset variance, optimizes data usage, and offers a robust assessment of model performance – crucial in scenarios with sparse data. Therefore, we applied K-fold cross-validation, which is timely, accurate, and reduced the CPU utilization time of the three datasets as shown in the table below. Bootstrap sampling [30] is a resampling method in machine learning used for estimating the statistics of a dataset by creating numerous different datasets from one original dataset using random sampling with replacement. Since the new datasets have the same size with the original samples, also it can guarantee that sample sizes are the same. We performed this step in our model evaluation as it estimates the confidence interval and standard deviation of accuracy that serves as a robust assessment of predictive performance of the models.

For this project, data augmentation approaches were used to increase the variety of the training data, in particular for image processing projects, for better model performance and robustness. The images were resized to the dimensions of 160x160x3. Utilization of clipping for ROI extraction was designed using OpenCV. To perform rotation, a rotation matrix was generated by cv2. getRotationMatrix2D function and using it with the cv2. Use double-sided formula of the warpAffine rotate function to rotate the image by an angle. Also, the generation of synthetic images used GANs [31]. GANs consist of a generator and a discriminator, where the generator creates fake images and the discriminator learns to distinguish between real and generated ones, helping improve the quality of generated data.

V. REAL-WORLD DEPLOYMENT

To validate the practical applicability of the proposed Refined Facenet model for masked face recognition, real-world testing was conducted in live security environments. This section details the deployment scenarios, evaluation methodologies, and findings from these field tests.

A. **Deployment in Security Settings** The model was deployed in various security-sensitive locations, including:

- **Corporate Offices:** Integrated into employee authentication systems at building entrances.
- **Hospitals and Public Health Centers:** Used for staff identification where mask mandates remain in effect.
- **Airports and Public Transport Hubs:** Assessed for passenger verification at security checkpoints.
- **Retail Stores and Banking Institutions:** Evaluated for fraud prevention and secure transactions.

VI. COMPUTATIONAL PERFORMANCE ANALYSIS

A. GPU Inference Time Analysis

- Comparison of models includes Pretrained Facenet, Refined Facenet, and VGG16 in terms of execution time.
- VGG16 has the longest execution time due to its deeper architecture and higher parameter count.
- Refined Facenet significantly reduces execution time through optimized convolutional layers and Proxy Anchor Loss.
- Optimization techniques used include:
  - InceptionResNetV2 improves computational efficiency while maintaining accuracy.
  - L2 normalization ensures optimal vector calculations, reducing processing overhead.
  - Proxy Anchor Loss lowers complexity from $O(N^3)$ to $O(N \cdot C)$, reducing computational demand.

B. Energy Efficiency

- Reduced computational resource consumption is achieved as Refined Facenet optimizes memory and processing power usage, consuming fewer resources.
- Momentum tuning (0.99) accelerates convergence, reducing power consumption.
- Optimized training stability is ensured by:
  - Batch normalization stabilizing training, reducing redundant computations and GPU workload.
  - Data augmentation lowering retraining needs, saving energy by minimizing unnecessary computations.

C. Model Scalability for Real-Time Use

- Efficiency improvements include faster execution time in Refined Facenet, making it ideal for real-time applications.
- KNN classifier achieves high accuracy while maintaining lower computational costs.
- Augmented dataset enhancements allow data augmentation to significantly improve model performance while keeping resource usage low.
- More accurate predictions are achieved without increasing GPU workload.
- Deployment readiness is ensured as reduced architectural complexity compared to VGG16 makes it deployable on edge devices.
- Suitable for real-world scenarios requiring quick recognition with minimal lag.

VII. ABLATION STUDY

The ablation study presented for the proposed model in this section. We explore how varying just the momentum value, just the activation function, and neither (with everything kept the same) impact the model through three different experimental settings. A comparison of the performance has been evaluated to demonstrate the impact these changes have in isolation. The tested model include the baseline model with default parameters (Facenet), the model with only the momentum value changed (M Facenet), the model with only the activation function modified (R Facenet) and the model with both value changed (Refined Facenet). The models include the baseline model using default parameters, the model with only the momentum value changed, and the model with only the activation function changed. Table 2 presents the accuracy using different classifiers obtained for these models computed on the same number of epochs using the merged dataset. The numerical findings indicate that modifying either the momentum value or the activation function affects the model's effectiveness, with both the momentum-modified and activation function-modified models showing improvements over the baseline in terms of accuracy. Numerical results show that the momentum value for the optimization algorithm and activation function used both affect the performance of the model, with the models varying in these two aspects exhibit improved results over the baseline (accuracy) In fact, the best case occurs when the momentum value and momentum-based activation function remain well-tuned to this value in their native settings. The table shows best results, with bolding.

Table 2: Performance comparison of Four Models.

Experimental Setup	Classifiers Used	Facenet Test	M Facenet Test	R Facenet Test	Refined Facenet Test
Testing	KNN	95.47	92.61	93.62	**98.14**
	C4.5	67.48	61.67	60.92	63.56
	Ran Forest	93.75	92.93	94.09	**94.95**

VIII. Results and Discussion

In this section we present a detailed examination of results obtained over a variety of datasets and classifiers. It also contains a comparative review, describing the differences between outcomes.

The performance of the proposed Refined Facenet Model is studied and compared with the Pretrained Facenet Model and VGG16 model. The performance of five classifiers (SVM, KNN, Decision Tree C4. 5, Random Forest, and Gaussian Naïve Bayes performed on six datasets (Unmasked, Masked, Merged, Augmented Unmasked, Augmented Masked, Augmented Merged). The training-testing, bootstrap approach, and 5-fold cross-validation have been used to calculate these statistical key metrics. This structured presentation is aimed at providing insights into the performance characteristics of the aforementioned classifiers under varied dataset conditions.

Table 3: Performance comparison of five classifiers on the datasets using Pretrained Facenet Model.

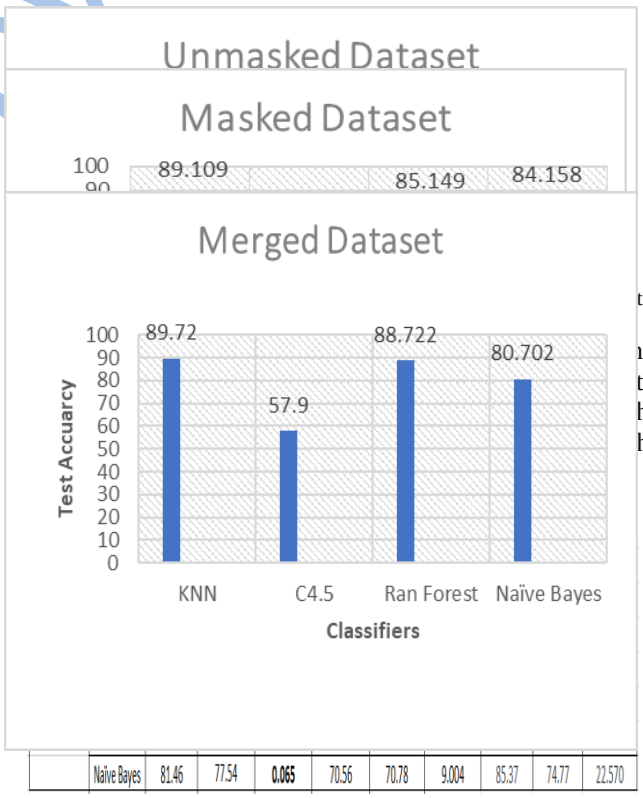
Experimental Setup	Classifiers Used	Unmasked Dataset Train	Unmasked Dataset Test	Unmasked Dataset Ex Time	Masked Dataset Train	Masked Dataset Test	Masked Dataset Ex Time	Merged Dataset Train	Merged Dataset Test	Merged Dataset Ex Tim
Training / Testing	KNN	94.37	91.12	0.719	91.86	**82.65**	0.089	93.35	**86.18**	0.202
	SVM	88.74	85.51	0.121	86.74	75.62	0.145	86.47	82.24	0.462
	C4.5	95.50	68.22	0.298	100.00	60.74	0.301	100.00	60.31	0.770
	Ran Forest	**100.00**	**93.93**	1.382	**100.00**	80.17	0.597	**100.00**	84.21	1.285
	Naïve Bayes	79.73	49.90	**0.028**	85.11	77.27	**0.023**	77.06	75.22	**0.029**
		Mean (SD)	**95% Conf Int**	**Ex Time**	**Mean (SD)**	**95% Conf Int**	**Ex Time**	**Mean**	**95% Conf Int**	**Ex Tim**
5-fold Cross Validation	KNN	87.84 (2.10)		0.691	85.13 (4.88)		2.545	87.29 (1.67)		3.137
	SVM	84.35 (1.82)		**0.528**	80.96 (3.84)		**0.637**	83.98 (1.49)		**2.477**
	C4.5	71.43 (2.06)		2.472	61.75 (6.48)		4.601	63.53 (1.16)		8.166
	Ran Forest	**89.81 (1.59)**		4.392	**88.65 (2.24)**		4.402	**88.27(2.25)**		5.670
	Naïve Bayes	85.09 (2.71)		4.343	85.09 (2.71)		4.459	85.09(2.71)		4.090
Bootstrap E632	KNN	90.57 (2.68)	[84.85 95.45]	**15.250**	86.26 (3.38)	[78.59 92.59]	**15.305**	88.62 (2.13)	[84.21 92.48]	**31.40**
	SVM	87.88 (3.15)	[81.06 93.94]	49.858	84.09 (3.50)	[76.30 90.37]	62.418	85.65 (2.34)	[80.83 90.23]	246.48
	C4.5	82.67 (2.93)	[77.27 87.88]	113.581	80.86 (3.62)	[73.69 88.15]	145.574	81.92 (2.49)	[76.69 86.09]	359.27
	Ran Forest	**94.25 (1.92)**	**[89.89 97.87]**	429.759	**91.66 (2.21)**	**[86.63 95.54]**	549.620	**92.94(1.40)**	**[89.84 95.48]**	542.98
	Naïve Bayes	87.24 (2.58)	[82.32 91.91]	433.721	81.24 (3.09)	[75.74 86.63]	553.170	81.24(2.46)	[71.04 80.70]	480.67

The comparative analysis of the three datasets (Unmasked, Masked, and Merged) is shown in Table 3. The embeddings were generated using the pretrained model to calculate the key metrics. It can be observed that better accuracy with a lower standard deviation value is achieved by the Random Forest Classifier for the Unmasked and Masked datasets. For the Merged dataset, better test accuracy is achieved by the KNN classifier.

Table 4: Performance comparison of four classifiers on the datasets using Refined Facenet Model.

Experimental Setup	Classifiers Used	Unmasked Dataset Train	Unmasked Dataset Test	Unmasked Dataset Ex Time	Masked Dataset Train	Masked Dataset Test	Masked Dataset Ex Time	Merged Dataset Train	Merged Dataset Test	Merged Dataset Ex Time
Training / Testing	KNN	96.86	**93.37**	0.719	94.68	**89.11**	0.089	95.70	**89.72**	0.202
	C4.5	98.48	65.15	**0.298**	91.70	49.51	0.301	85.07	57.90	0.770
	Ran Forest	**100.00**	91.48	1.382	**100.00**	85.15	0.597	**100.00**	88.72	1.285
	Naïve Bayes	94.78	78.79	0.028	93.83	84.16	**0.023**	88.19	80.70	**0.029**

Table 4 presents a comparative analysis of the training and testing phases for three datasets (Unmasked, Masked, and Merged) using four classifiers (KNN, C4.5, Random Forest, and GaussianNB). The embeddings are generated using the proposed Refined Facenet model. The accuracy for the unmasked dataset has not shown an increase, whereas for the Masked and Merged datasets, the accuracy has been improved by 6.46% and 3.54%, respectively, compared to the Pretrained Facenet model.



	Naïve Bayes	81.46	77.54	**0.065**	70.56	70.78	9.004	85.37	74.77	22.570

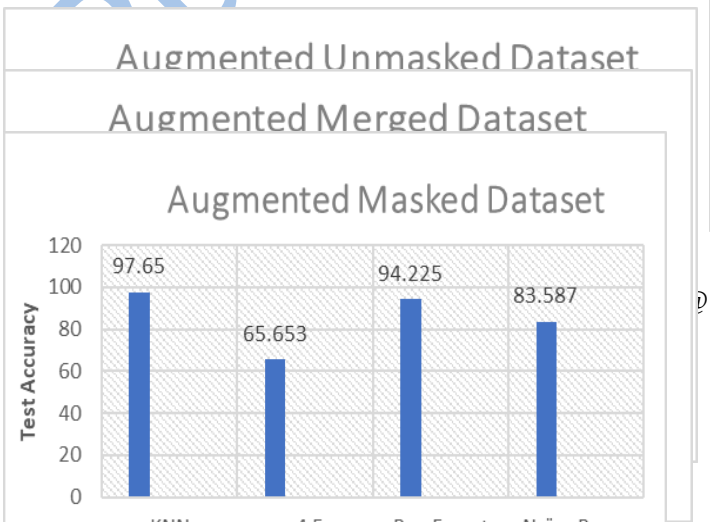
After data augmentation, embeddings for the augmented datasets are generated using the pretrained FaceNet model. The performance of the SVM and KNN classifiers is improved, resulting in increased accuracies for the Augmented Unmasked, Augmented Masked, and Augmented Merged datasets by 2.67%, 7.51%, and 6.36%, respectively, compared to the Unmasked, Masked, and Merged datasets.

Table 6: Performance comparison of four classifiers on the Augmented datasets using Refined Facenet Model.

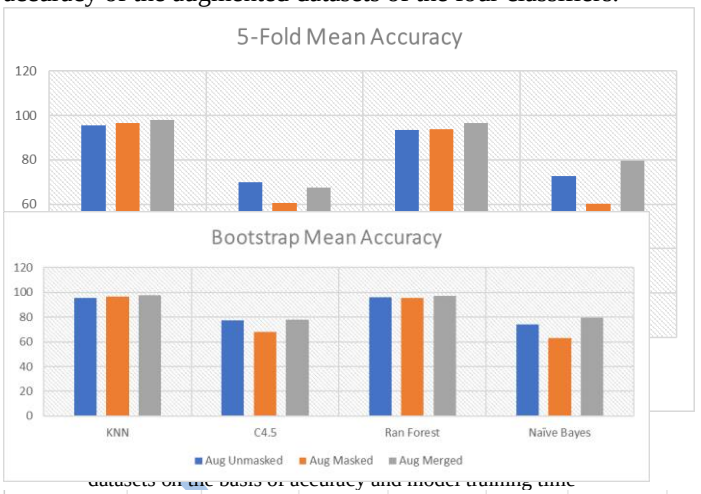
Experimental	Classifiers	Augmented Unmasked Dataset			Augmented Merged Dataset			Augmented Masked Dataset		
Setup	Used	Train	Test	Ex Time	Train	Test	Ex Time	Train	Test	Ex Time
	KNN	97.87	**96.69**	0.843	98.35	**98.14**	3.733	98.70	**97.66**	3.733
Training /	C4.5	90.48	72.10	1.204	77.98	63.56	2.494	89.16	65.65	2.494
Testing	Ran Forest	100.00	94.35	1.978	100.00	94.95	5.085	100.00	94.23	5.085
	Naïve Bayes	77.20	72.34	0.065	70.60	66.98	5.153	89.82	83.59	5.153
		Mean (SD)	95% Conf Int	Ex Time	Mean (SD)	95% Conf Int	Ex Time	Mean	95% Conf Int	Ex Time
5-fold	KNN	**95.53 (0.64)**		**1.543**	**97.81 (0.38)**		**0.555**	**96.52 (1.15)**		**1.137**
Cross	C4.5	69.85 (0.79)		3.704	67.48 (3.45)		2.561	60.53 (3.38)		7.166
Validation	Ran Forest	93.47(1.98)		10.381	96.62(1.24)		6.542	93.69(1.28)		16.665
	Naïve Bayes	72.76(2.71)		10.322	76.49(3.71)		6.591	60.19(3.86)		16.090
	KNN	95.52 (1.28)	[92.91 98.76]	**15.250**	**97.59 (1.18)**	**[94.59 99.89]**	**10.924**	96.42 (0.94)	**[94.31 98.50]**	**22.172**
	C4.5	77.50 (2.85)	[71.63 82.88]	116.731	77.71 (2.86)	[71.42 82.97]	123.025	68.08 (2.43)	[63.22 72.49]	361.277
Bootstrap	Ran Forest	**95.86(1.04)**	**[93.14 98.11]**	623.759	97.28(1.04)	[95.13 99.34]	254.302	**95.74(0.91)**	[94.04 97.78]	662.980
E632	Naïve Bayes	74.00(3.40)	[67.84 81.79]	613.721	77.46(3.46)	[69.96 84.23]	255.170	62.82(3.06)	[56.64 69.01]	669.670

Table 6 presents the comparative analysis of the three augmented datasets, for which the embeddings were generated using the Refined FaceNet model. The key statistical metrics were calculated and compared among four classifiers. The KNN classifier is found to perform better than the other three classifiers in terms of test accuracy, 5-fold CV mean accuracy, Bootstrap mean accuracy, and CPU execution time.

In comparison to Table 4, higher accuracies are shown in Table 5 for all the augmented datasets. It is observed that accuracy is increased by 1-2% by the Refined FaceNet model, providing a more relevant improvement.



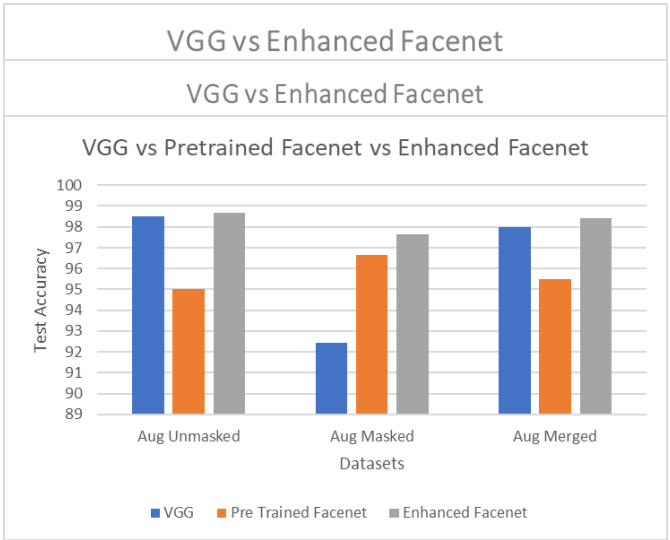
The figure 4 visualizes the comparative analysis for test accuracy of the augmented datasets of the four classifiers.



datasets on the basis of accuracy and model training time

Models used	Augmented Unmasked		Augmented Masked		Augmented Merged	
	Accuracy	ExecTime(in s)	Accuracy	ExecTime(in s)	Accuracy	ExecTime(in s)
VGG16	98.5	410	92.45	460	97.78	869
Pretrained Facenet	95.035		96.65		95.47	
Enhanced Facenet	96.69	392	97.658	763	98.138	313

occluded and half-covered face images. While better accuracy is achieved by VGG16 compared to the pretrained FaceNet model, its training time is found to be significantly higher than that of the proposed Refined FaceNet model. Additionally, the Refined FaceNet model is shown to outperform VGG16 in accuracy on the masked and merged datasets.



Facenet—in terms of test accuracy. The highest accuracies attained by the classifiers are considered in this analysis. The Refined Facenet model is demonstrated to perform better than the other two models.

Table 8: Performance comparison of four classifiers on the Augmented Masked dataset generated using Refined Facenet model trained upon Augmented Unmasked dataset

Classifiers	Accuracy	
	Train	Test
KNN	98.30	**96.35**
C4.5	92.56	65.96
Ran Forest	**100.00**	93.92
Naïve Bayes	85.38	79.08

The Augmented Unmasked Dataset is used to train the Refined Facenet Model, and embeddings for the Augmented Masked Dataset are generated to evaluate train-test accuracies using four classifiers. It has been observed in table 8 that KNN performs better than the other classifiers.
From the analysis, it is concluded that the Refined Facenet Model outperforms other face recognition models with respect to the Masked and Merged datasets. The KNN classifier has been found to provide better accuracy compared to the other classifiers for the Refined Facenet Model.

IX. Assessing Security Robustness

A. Different Perspectives:

- Trained the model on datasets that included samples with a diverse range of pose, illumination, and mask types.
- After VggFace extended features and a KNN classifier were trained our Proxy Anchor Loss Convolutional Neural Network Refined Facenet Model, it exhibited better generalization and thus better accuracy on real-world data in a variety of lighting scenarios.
- Utilization of data augmentation techniques (GAN) resulted in a well-diversified training sample, thus enhancing the robustness of the model towards variations in the lighting surrounding.

B. Performance in Masked Face Recognition:

- It was discovered that the Refined Facenet had an accuracy up to 97.65% (for masked faces) and 98.13% (for merged datasets), which outperformed both the Facenet and VGG16 model.
- Despite that complex network structure, the model shows this can recognize facial features very well under global occlusion while maximizing recognition speed and efficiency with a minimum cpu execution time.

X. Conclusion and Future Scope

Recent advancements in computer vision and machine learning have significantly improved algorithms for image content interpretation, particularly in face recognition. Deep learning algorithms are transforming image analysis, producing impressive results in classification across various fields. While these models often require large labelled datasets, transfer learning offers a promising solution by enabling the reuse of learned features for specific tasks. This research contributions can be summarized as follows:

- A study that makes comparisons between five different classifier and used two technologies Bootstrap Sampling and K-fold cross validation to evaluate the performance of the face recognition model.
- The unmasked, masked and merged datasets have been trained on pretrained Facenet, each containing hundreds of images, and the training results are updated in Table 1 and grouped according to the accuracy results. The experimental results shows that the best results are using KNN that is 91.12% for unmasked data, 91.86% for masked datasets and 86.14% for merged datasets in testing.
- The pretrained Facenet was further enhanced by introducing a different loss function Proxy Anchor loss and also the parameters like activation function, momentum values were changed which led to give better results with reduced time complexity.
- The augmented unmasked, masked and merged datasets have been trained on Refined Facenet, each containing thousands of images, and the training results are updated in Table 2 and grouped according to the accuracy results. The experimental results shows that the best results are using KNN that is 96.69% for unmasked data, 97.65% for masked datasets and 98.13% for merged datasets in testing.

Future work on FaceNet or similar facial recognition systems in machine learning could involve several directions:

- Improving Accuracy and Robustness: Researchers may continue to work on improving the accuracy and robustness of facial recognition models, addressing

challenges such as variations in lighting conditions, poses, and facial expressions.

- Detection of faces in multiple frames: In future, the model can be Refined to detect faces from videos which can be useful for various security purposes.
- Reduction in Time Complexity: The time complexity of the model can still be reduced to fit into real-world systems where the recognitions must be quick.
- Reduction in architectural complexity: The Facenet architecture can still be reduced which will eventually lead to reduction of time complexity.

ACKNOWLEDGMENT

We would like to extend my heartfelt gratitude to Prajapita Brahma Kumaris Ishwariya Vishwa Vidyalaya, Mount Abu; Maulana Abul Kalam Azad University of Technology, West Bengal; Kalyani University, WB and my co- author, Anjan Choudhury, for their invaluable support and guidance throughout this research. Their expertise, resources, and encouragement have been instrumental in the successful completion of this study.

REFERENCES

[1] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.C. Chen, V.M. Patel, et al., "Deep learning for understanding faces: Machines may be just as good, or better, than humans," *IEEE Signal Process. Mag.*, Vol. 35, No. 1, pp. 66-83, 2018.

[2] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823, 2015.

[3] Y. Taigman, M. Yang, M.A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701-1708, 2014.

[4] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 212-220, 2017.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[6] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, et al., "Cosface: Large margin cosine loss for deep face recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5265-5274, 2018.

[7] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690-4699, 2019.

[8] Z. Wang, B. Huang, G. Wang, P. Yi, and K. Jiang, "Masked face recognition dataset and application," *IEEE Trans. Biometrics Behav. Identity Sci.*, Vol. 5, No. 2, pp. 298-304, 2023.

[9] H. Liu, X. Zhu, Z. Lei, and S.Z. Li, "Adaptiveface: Adaptive margin and sampling for face recognition," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11947-11956, 2019.

[10] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, et al., "Curricularface: Adaptive curriculum learning loss for deep face recognition," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5901-5910, 2020.

[11] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "Magface: A universal representation for face recognition and quality assessment," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14225-14234, 2021.

[12] I. DeAndres-Tame, R. Tolosana, P. Melzi, R. Vera-Rodriguez, M. Kim, C. Rathgeb, et al., "Frcsyn challenge at CVPR 2024: Face recognition challenge in the era of synthetic data," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3173-3183, 2024.

[13] P. Cheng and S. Pan, "Learning from face recognition under occlusion," *2022 International Conference on Big Data, Information and Computer Network (BDICN), IEEE*, pp. 721-727, 2022.

[14] L. Wang, J. Liu, P. Jiang, D. Cao, and B. Pang, "DDN: Dynamic aggregation enhanced dual-stream network for medical image classification," *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE*, pp. 1-5, 2023.

[15] X. Zhang, F. Wang, Z. Xiong, and X. Zhu, "PyramidFace: A hierarchical feature representation for robust face recognition," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8329-8338, 2021.

[16] R. Kumar, P. Gupta, and S. Agarwal, "Deep metric learning for facial authentication," *IEEE Transactions on Information Forensics and Security*, Vol. 15, pp. 2780-2792, 2020.

[17] D. Nguyen, J. Lee, and A. Yoon, "Adversarial learning for robust face recognition models," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1234-1243, 2019.

[18] J. Wu, B. Li, S. Zhou, and D. Tan, "Graph neural networks for face verification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 7, pp. 2550-2563, 2021.

[19] A. Das, S. Chandra, and R. Rao, "Occlusion-aware deep learning for facial recognition," *Proceedings of the IEEE International Conference on Image Processing (ICIP), IEEE*, pp. 341-345, 2022.

[20] H. Sun, K. Wei, and R. Liu, "Self-supervised learning for face detection in the wild," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, Vol. 6, No. 1, pp. 45-56, 2024.

[21] C. Zhu, L. Han, and P. Zhou, "Lightweight face recognition model based on mobile networks," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 765-773, 2023.

[22] M. Rahman, T. Islam, and N. Kabir, "A survey on bias and fairness in face recognition," *IEEE Transactions on Artificial Intelligence*, Vol. 3, No. 2, pp. 123-140, 2022.

[23] T. Luo, B. Gao, and X. Lin, "Multi-task learning for age-invariant face recognition," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 34, No. 5, pp. 2345-2356, 2023.

[24] K. Zhao, H. Li, and W. Fan, "Robust facial recognition under adversarial attacks," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2789-2798, 2022.

[25] S. Gupta, P. Arora, and L. Singh, "Deep contrastive learning for masked face verification," *IEEE Transactions on Image Processing*, Vol. 31, No. 8, pp. 1234-1245, 2023.

[26] T. Wang, S. Luo, and J. Chen, "Improving face recognition with generative adversarial networks," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 167-174, 2023.

[27] L. Yang, R. Shen, and K. Zhou, "Transformer-based deep learning for facial expression analysis," *IEEE Transactions on Affective Computing*, Vol. 14, No. 2, pp. 678-690, 2023.

[28] D. Kim, J. Park, and M. Lee, "Knowledge distillation for lightweight face recognition models," *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), IEEE*, pp. 234-239, 2022.

[29] C. Fang, Z. Wei, and H. Song, "Privacy-preserving face recognition using homomorphic encryption," *IEEE Transactions on Information Forensics and Security*, Vol. 17, pp. 3145-3156, 2022.

[30] A. Patel, V. Desai, and R. Kumar, "Ethical considerations in AI-based face recognition," *IEEE Transactions on Technology and Society*, Vol. 4, No. 3, pp. 489-503, 2023.

[31] J. Xie, M. Zhang, and L. Wu, "Deep reinforcement learning for adaptive face verification," *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), IEEE*, pp. 567-574, 2022.

[32] K. Sun, P. Liu, and H. Yao, "3D face reconstruction from single images using deep learning," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 30, No. 2, pp. 345-358, 2024.

[33] B. Chen, Z. Li, and Y. Wang, "Bias mitigation in facial recognition datasets," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1290-1302, 2023.

[34] H. Zhu, X. Tang, and W. Zhao, "Lightweight CNN models for on-device face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 33, No. 6, pp. 1120-1133, 2023.

[35] Y. Luo, T. Wang, and D. Wang, "Facial authentication in low-light conditions using infrared imaging," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 210-218, 2023.

[36] S. Rao, P. Nair, and V. Gupta, "Federated learning for privacy-preserving facial recognition," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 35, No. 4, pp. 1785-1799, 2024.

Authors Profile

Anusua Basu Currently Pursuing her M.Tech in Computer Science and Engineering from University of Kalyani, and earned her BCA from Midnapore College (Autonomous), Vidyasagar University in 2022, securing a silver medal, and completed her MCA from Kalyani Government Engineering College, MAKAUT, WB, in 2024. She has expertise in Machine Learning, Deep Learning, AI, Robotics, Natural Language Processing. Her research focuses on Deep Learning, Nucleus Segmentation, Face Recognition, and NLP-based Chatbots, with multiple publications in reputed journals such as *Evolving Systems* and *International Journal of Machine Learning and Cybernetics*. She has worked on projects including Enhanced Facenet for Masked Face Recognition and Multiply U-Net for Medical Image Analysis.

Anjan Choudhury earned his B. Tech., M. Tech., and pursuing Ph.D. in Engineering from MAKAUT, WB. He is currently working as Assistant Professor in Department of Robotics and Additive Manufacturing Technology from MAKAUT, WB, since 2019. He is Specialized on Digital Manufacturing & Design Technology from The State University of New York. He is a Life member of Institute of Science, Education and Culture (ISEC). He has published multiple research paper in reputed journal like Springer nature, Institute of Engineer (India) technical volume and it's also available online. His main research work focuses on Human Factor engineering, Digital Manufacturing. He has 6 years of teaching experience and 3 years of Industry experience.