VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Belagavi, Karnataka-590018, Karnataka

MINI PROJECT REPORT

ON

"GENDER DETECTION USING FACIAL THERMAL IMAGES"

Submitted in partial fulfilment of the requirements for the course

Mini Project (21CSMP67)

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Pinni Raga Sruthi	**1JS21CS102**
Rakshith V Patil	**1JS21CS113**
Raunak Singh Rathour	**1JS21CS115**
Sahithi Srujana C	**1JS21CS122**

Under the Guidance of

Ms K. V. Shanthala

Assistant Professor, Dept of CSE,

JSSATE, Bengaluru

JSS ACADEMY OF TECHNICAL EDUCATION, BENGALURU

Department of Computer Science and Engineering

2023-2024

JSS ACADEMY OF TECHNICAL EDUCATION

JSS Campus, Dr.Vishnuvardhan Road, Bengaluru-560060

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that **A MINI PROJECT REPORT entitled "GENDER DETECTION USING FACIAL THERMAL IMAGES"** has successfully carried out by **Pinni Raga Sruthi (1JS21CS102) , Rakshith V Patil (1JS21CS113), Raunak Singh Rathour (1JS21CS115), Sahithi Srujana C (1JS21CS122)** in partial fulfilment for the course Mini Project (21CSMP67) of 6^{th} Semester **Bachelor of Engineering in Computer Science and Engineering** in **Visvesvaraya Technological University Belagavi** during the year 2023-2024. It is certified that all corrections and suggestions indicated for Internal Assessment have been incorporated in the report deposited in the department library. The mini project report has been approved as it satisfies the academic requirement in respect of the project work prescribed for the said degree.

Ms K. V. Shanthala	**Dr. Mallikarjuna P B**
Asst. Professor,	**Professor & Head,**
Dept. of CSE,	Dept. of CSE,
JSSATE, Bengaluru	JSSATE, Bengaluru

ABSTRACT

This project explores the enhancement of gender classification systems by integrating facial thermal images, leveraging their invariance to lighting conditions and ability to capture unique physiological differences between genders. The methodology involves utilizing pretrained convolutional neural networks (CNNs), particularly ResNet-50, which is fine-tuned for the gender classification task. By modifying the final layers of the ResNet-50 model to adapt to the new task and employing advanced deep learning techniques, the system demonstrates improved accuracy and reliability. This approach not only advances the field of gender classification but also provides a robust solution for diverse real-world applications, setting a new benchmark in the domain.

ACKNOWLEDGEMENT

We express our humble pranams to his holiness **Jagadguru Sri Sri Sri Shivaratri Deshikendra Mahaswamiji** for showering his blessings on us to receive good education and have a successful career.

We express our sincere thanks to our beloved principal, **Dr. Bhimasen Soragaon** for having supported us in all our academic endeavors.

We are also forever grateful to **Dr. P.B. Mallikarjuna**, Head of the Department, Computer Science and Engineering, for his unending support, guidance and encouragement in all our ventures.

We are thankful for the resourceful guidance, timely assistance and graceful gesture of our guide **Ms. Shanthala K V,** Assistant professor, Department of Computer Science and Engineering, who helped us in every aspect of our project work.

The completion of any project involves the efforts of many people. We have been lucky enough to have received a lot of support from all ends during the course of this project. So, we take this opportunity to express our gratitude to all whose guidance and encouragement helped us emerge successful.

Last but not the least; we would be immensely pleased to express our heartfelt thanks to all the teaching and non-teaching staff of the department of CSE and our friends for their timely help, support and guidance.

Pinni Raga Sruthi (1JS21CS102)
Rakshith V Patil (1JS21CS113)
Raunak Singh Rathour (1JS21CS115)
Sahithi Srujana C (1JS21CS122)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The project focuses on leveraging uncooled thermal imaging for gender classification, aiming to overcome challenges associated with visible spectrum images, such as varying illumination conditions and occlusions. By utilizing thermal facial images, the system ensures reliability in diverse conditions. Various state-of-the-art convolutional neural networks, including ResNet-50, DenseNet-121, and ResNet-101, are fine-tuned using transfer learning on the Tufts University thermal facial image dataset. This approach not only enhances the accuracy of gender classification but also demonstrates the potential of thermal imaging as a complementary modality for applications in security, human-computer interaction, and automotive systems.

1.1 Motivation

The motivation for this project arises from the limitations of traditional gender classification methods using visible spectrum images, which are often affected by illumination, shadows, and varying lighting conditions. To overcome these challenges, the study explores the use of thermal imaging, which captures infrared radiation and offers consistent performance regardless of lighting. By leveraging advanced deep learning models such as ResNet-50, DenseNet-121, and ResNet-101, this research aims to develop a more reliable and accurate gender classification system using thermal facial images.

1.2 Existing System Overview

Conventional Machine Learning-Based Methods

The existing system for gender classification employs Support Vector Machines (SVM) to improve performance using histogram-based features on Near-Infrared (NIR) and thermal spectra. Feature extraction techniques such as Histogram of Oriented Gradients (HoG) and Multi-Scale Local Binary Patterns (MLBP) are utilized for joint visible and thermal spectrum data, enhancing classification accuracy. Multimodal datasets, including audiovisual, thermal, and physiological recordings, are employed to extract feature values for automatic classification. Various conventional machine learning algorithms, such as SVM, Linear Discriminant Analysis (LDA), and Adaboost, are utilized for gender representation and recognition based on facial images, leveraging the complementary strengths of different spectral data.

Deep Learning-Based Methods

Deep CNN structures are widely used for applications requiring precise accuracy levels, such as medical image analysis and surveillance systems. Pretrained models listed by Canziani et al. can be used for gender recognition and other applications. In this project, ResNet-50, DenseNet-121, and ResNet-101 are evaluated for their performance in gender classification tasks using thermal facial images. ResNet-50's 50-layer architecture effectively mitigates the vanishing gradient problem, DenseNet-121's dense connectivity pattern enhances feature reuse, and ResNet-101's deeper 101-layer architecture enables the learning of more complex features. By cross-validating these models on a validation set of thermal images acquired using a prototype uncooled thermal camera, we aim to determine the most suitable model for accurate gender classification in thermal imagery.

1.3 Proposed System Overview

The proposed system focuses on the performance estimation of various state-of-the-art Convolutional Neural Networks (CNNs) for gender classification using thermal facial images. Specifically, it evaluates the effectiveness of ResNet-50, DenseNet-121, and ResNet-101 models, each trained for 50 epochs. The CNNs were first pretrained on ImageNet to harness their generalization capabilities. They were then fine-tuned on thermal facial data to assess their effectiveness. The system highlights the potential of using thermal imaging combined with deep learning models for effective gender classification.

Data acquisition involved using a prototype thermal VGA camera to capture facial images in an indoor lab setting with varied poses for diversity. The data were processed into image sequences and split into training and testing sets for cross-validation. ResNet-50, DenseNet-121, and ResNet-101 were trained on a combined dataset of Carl and locally gathered thermal images for 50 epochs using optimizers like SGD and Adam. Model performance was evaluated using accuracy metrics based on true positives, false positives, true negatives, and false negatives. The results include training, validation, and test accuracy, as well as the total number of parameters for each model.

CHAPTER 2

LITERATURE SURVEY

2.1 Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation

Authors: Philip Smith and Cuixian Chen

Summary: This paper explores the application of transfer learning using deep convolutional neural networks (CNNs) for gender recognition and age estimation from images. The authors utilized VGG19 and VGGFace models with pre-trained weights, experimenting with various design and training techniques to enhance prediction accuracy. Notable methods included input standardization, data augmentation, and label distribution age encoding.

The proposed hierarchical model, which first classifies gender and then predicts age using gender-specific models, achieved a gender recognition accuracy of 98.7% and a mean absolute error (MAE) of 4.1 years in age estimation.

2.2 Thermographic Imaging of Facial Skin—Gender Differences and Temperature Changes Over Time in Healthy Subjects

Authors: J. Christensen, M. Væth, and A. Wenzel

Journal: Dentomaxillofacial Radiology

Summary: This study investigates the autonomic regulation of facial temperature in response to stress using InfraRed Thermography (IRT). It reveals that facial temperature is significantly correlated with physiological indicators such as heart rate variability (HRV) and electrodermal activity (EDA), especially noting that the nose temperature is more strongly linked to EDA and heart rate variability under stress conditions.

The study highlights gender differences in the coupling between autonomic responses and thermal dynamics, underscoring the complexity of physiological processes influencing facial temperature during emotional stimuli. The findings suggest that thermal imaging may serve as a viable tool for understanding the physiological underpinnings of emotional states.

2.3 Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces

Authors: Makinen and R. Raisamo

Journal: IEEE Transactions on Pattern Analysis and Machine Intelligence

Summary: This study investigates various approaches to gender classification using facial images that have been automatically detected and aligned. The authors evaluate several gender classification methods, including both traditional and modern techniques, and provide a comparative analysis of their performance.

The study demonstrates that face alignment significantly improves the accuracy of gender classification by reducing variability due to pose, lighting, and other factors. The findings emphasize the importance of face detection and alignment in enhancing classification accuracy, with potential applications in human-computer interaction, security, and demographic analysis.

2.4 Multimodal Camera-Based Gender Recognition Using Human-Body Image with Two-Step Reconstruction Network

Authors: N. R. Baek et al.

Journal: IEEE Access

Summary: The paper introduces a comprehensive method for gender recognition using human-body images captured from a distance. The authors address the limitations of traditional approaches that rely on high-resolution facial images, especially in surveillance environments. The proposed method employs a two-step reconstruction network utilizing convolutional neural networks (CNNs) for image denoising and super-resolution.

By combining visible-light and infrared (IR) images, the method improves recognition accuracy and handles challenges such as motion blur and illumination variations. The study demonstrates that this approach significantly enhances gender recognition.

performance and is feasible for real-time applications in intelligent surveillance and autonomous vehicles.

2.5 Densely Connected Convolutional Networks

Authors: G. Huang et al.

Journal: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

Summary: Dense Convolutional Networks (DenseNets) introduce a novel architecture that connects each layer to every other layer in a feed-forward manner. This dense connectivity pattern alleviates the vanishing-gradient problem, enhances feature propagation, encourages feature reuse, and significantly reduces the number of parameters compared to traditional convolutional networks.

DenseNets have been evaluated on competitive object recognition tasks, achieving state-of-the-art results while requiring fewer parameters and less computation. For instance, DenseNet-BC with 190 layers achieved error rates of 3.46% on CIFAR-10+ and 17.18% on CIFAR-100+, outperforming existing architectures. DenseNets exhibit efficient parameter usage, with DenseNet-BC requiring about one-third the parameters of ResNets to achieve similar accuracy.

2.6 Deep Residual Learning for Image Recognition

Authors: K. He et al.

Journal: IEEE Conference on Computer Vision and Pattern Recognition

Summary: The paper introduces a deep residual learning framework to facilitate the training of very deep neural networks, addressing the degradation problem where deeper networks exhibit higher training errors and saturation in accuracy. By reformulating the learning process to focus on residual functions, the authors propose that it is easier to optimize these residual mappings rather than the original unreferenced functions. The residual learning principle is shown to be applicable across various visual recognition tasks, leading to significant improvements in accuracy with increased depth. A 152-layer ResNet achieved a top-5 error rate of 3.57% on ImageNet and first-place finishes in several other competitions, highlighting its generalization capability and efficiency for training.

CHAPTER 3

SYSTEM ANALYSIS

3.1 Hardware and Software Requirements

Hardware Specifications:

To efficiently process and analyze the Tufts Face Dataset and train deep learning models like ResNet50, ResNet101, and DenseNet-201, robust hardware specifications are necessary. The following hardware was utilized:

CPU: Intel i5 11th Gen Processor, providing substantial computational power for data preprocessing tasks.

GPU: NVIDIA GTX 1080 GPU with 8GB VRAM, facilitating accelerated training of deep neural networks.

RAM: 16GB DDR4 RAM, ensuring smooth handling of large datasets and preventing bottlenecks during data loading.

Storage: 200GB SSD for fast read/write speeds and efficient storage of the dataset and model checkpoints.

Software Tools and Libraries:

PyTorch: PyTorch is a powerful and flexible deep learning framework widely used for developing and training neural networks. It provides support for dynamic computational graphs, which makes it easier to debug and experiment with. PyTorch includes a variety of prebuilt functions for tensor operations, neural network layers, and optimization algorithms, making it a comprehensive tool for machine learning and deep learning projects.

Torchvision: Torchvision is a library that complements PyTorch by providing tools specifically designed for computer vision tasks. It includes popular datasets, model architectures, and image transformation utilities. Torchvision makes it easier to preprocess image data, load datasets, and implement state-of-the-art neural networks for image classification, object detection, and other vision-related applications.

Datasets: This module in Torchvision includes several popular datasets like CIFAR-10, ImageNet, and MNIST. It provides a standardized way to download, preprocess, and load these datasets for training and evaluating models. This helps streamline the workflow by abstracting the data handling processes.

Models: The models module offers a collection of pretrained models for various vision tasks. These models, such as ResNet, VGG, and DenseNet, come pre-trained on large datasets like ImageNet. Using pretrained models allows for transfer learning, where you can fine-tune the model on your specific dataset, saving time and computational resources.

Transforms: The transforms module includes a variety of image transformation functions that are used for preprocessing data. Common transformations include resizing, cropping, normalization, and augmentation techniques like random rotation and flipping. These transformations help improve the robustness and generalization of the trained models.

torch.nn: This module provides classes and functions to build neural networks. It includes layers such as convolutional layers, pooling layers, activation functions, and fully connected layers. The module also includes loss functions like CrossEntropyLoss and MeanSquaredError, which are essential for training neural networks.

torch.optim: The optim module offers a variety of optimization algorithms used to train neural networks. Examples include SGD (Stochastic Gradient Descent), Adam, and RMSprop. These optimizers adjust the model parameters to minimize the loss function during training, improving the model's performance on the given task.

DataLoader: This class provides an efficient way to load data in batches, shuffle the data, and use multiple worker processes for loading data. It is essential for handling large datasets that cannot be loaded into memory all at once. DataLoader makes the training process more efficient by enabling parallel data loading and batching.

time: The time module is a standard Python library used to measure the time taken for various operations. In the context of training neural networks, it is often used to benchmark the training process, measure the duration of each epoch, and optimize the overall training pipeline.

torchsummary: This library provides a convenient way to visualize the architecture of a neural network model. The summary function displays the model's layers, the shape of the input and output at each layer, and the number of parameters. This helps in understanding the complexity and size of the model.

lr_scheduler: The lr_scheduler module in PyTorch provides classes for learning rate scheduling, which adjusts the learning rate during training according to a predefined schedule. Common schedulers include StepLR, MultiStepLR, and ReduceLROnPlateau. Using learning rate schedulers can improve model convergence and prevent overfitting.

NumPy: NumPy is a fundamental library for numerical computing in Python. It provides support for large multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays. In deep learning, NumPy is often used for data manipulation, mathematical operations, and handling arrays before converting them into PyTorch tensors.

Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. It is widely used for plotting graphs and visualizing data. In the context of neural networks, Matplotlib is often used to visualize training and validation loss, accuracy over epochs, and to display sample images from the dataset.

os: The os module provides a way to interact with the operating system in Python. It is used for tasks such as file and directory manipulation, reading environment variables, and running system commands. In deep learning projects, it is often used to handle file paths, create directories for saving models, and manage dataset files.

PIL (Python Imaging Library): The PIL library, now maintained under the name Pillow, is used for opening, manipulating, and saving image files. It supports a wide range of image formats and provides various image processing functions. In this project, PIL is used to handle image loading and preprocessing before converting them into PyTorch tensors for model training.

3.2 Data Flow Diagram

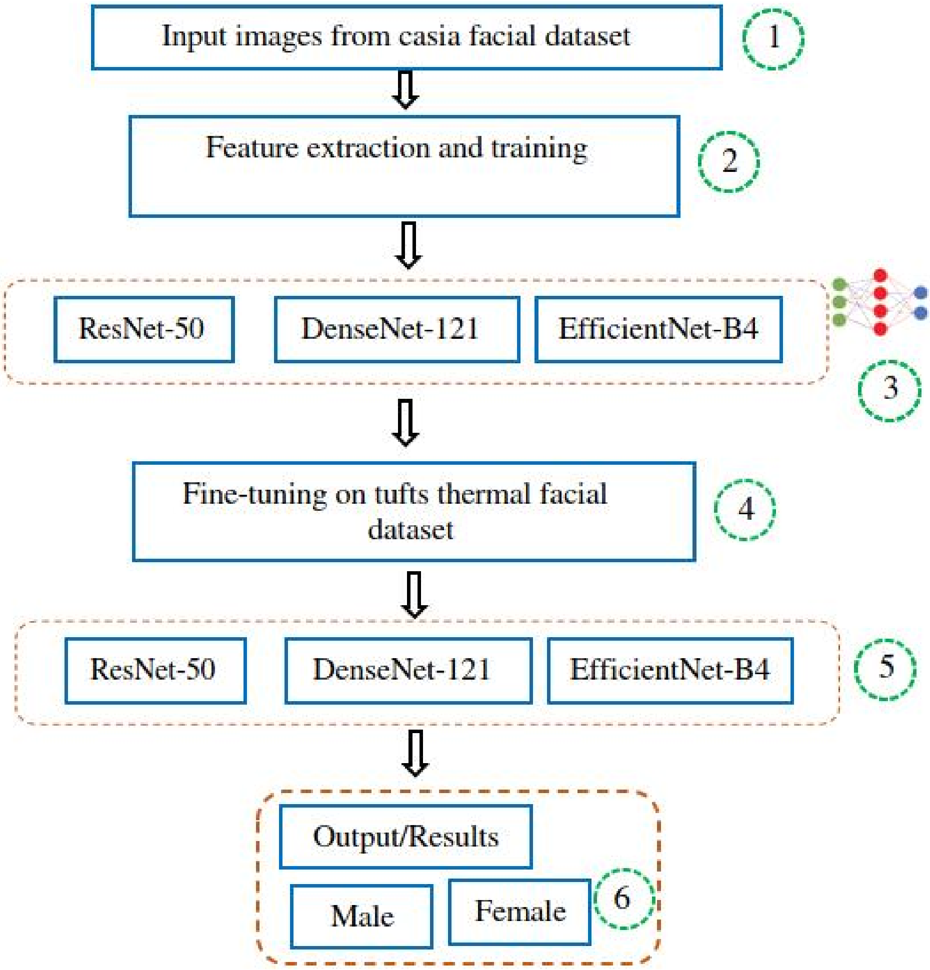

Fig 3.1 Data Flow Diagram

The high-level system architecture outlines the major components and their interactions within the project. It includes data collection, preprocessing, model training, evaluation, and deployment stages:

1. **Input Images from CASIA Facial Dataset**: The system starts by collecting input images from the CASIA facial dataset, which provides a diverse set of facial images for initial training.
2. **Feature Extraction and Training**: The collected images undergo feature extraction and training using the selected pretrained models (ResNet-50, DenseNet-121, and EfficientNet-B4). This step involves leveraging the models' ability to learn from complex features present in the images.

3. **Model Selection**: ResNet-50, DenseNet-121, and EfficientNet-B4 are chosen for their proven effectiveness in various image classification tasks. These models are pretrained on large datasets, providing a strong starting point for transfer learning.
4. **Fine-Tuning on Tufts Thermal Facial Dataset**: The pretrained models are fine-tuned on the Tufts thermal facial dataset, which contains thermal images specifically captured for this project. This fine-tuning helps the models adapt to the specific characteristics of thermal images.
5. **Evaluation and Output**: The fine-tuned models are evaluated on a validation set, and their performance is assessed based on their ability to classify gender. The final output consists of the predicted gender (male or female) for each image.

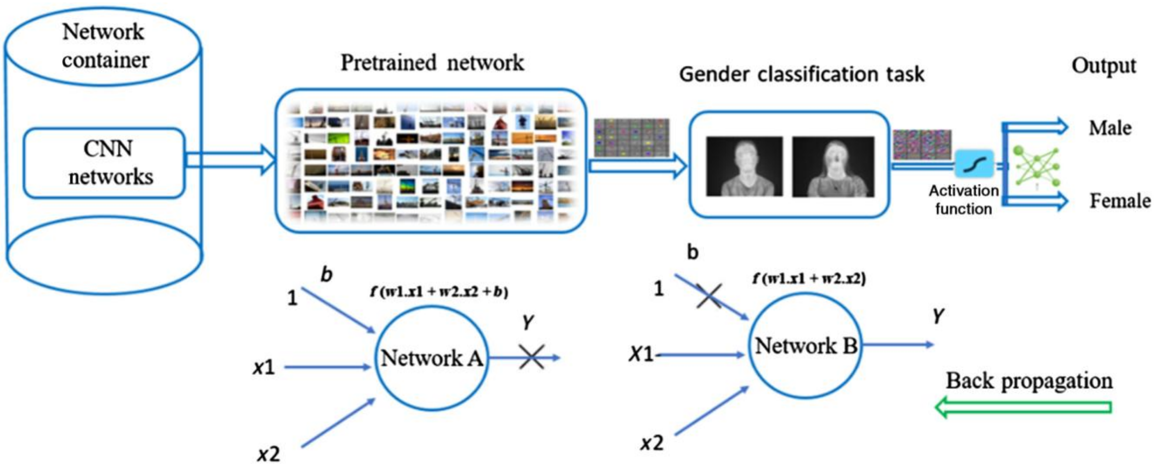

Fig 3. 2 CNN training structure

The diagram illustrates the process of using pretrained convolutional neural networks (CNNs) for the gender classification task based on facial thermal images. The "Network container" consists of various CNN architectures that have been pretrained on a diverse set of images. These pretrained networks, which include models like ResNet-50, ResNet-101, and DenseNet-121, are employed as the foundation for the gender classification system. These networks are capable of extracting rich features from input images due to their extensive training on large-scale image datasets.

In the next stage, these pretrained networks are fine-tuned for the specific gender classification task. This involves modifying the final layers of the networks to adapt them to the new task of identifying gender from facial thermal images. The input to these fine-tuned networks consists of facial thermal images, which are processed to extract relevant features that distinguish male from female faces. The extracted features are then passed through an activation function to produce a classification output.

The diagram also highlights the backpropagation process, essential for training neural networks. Network A represents the initial state where the input features (x1, x2) are weighted and combined to produce an output. Network B shows the updated state after backpropagation, where the weights (w1, w2) and biases (b) have been adjusted to minimize the classification error. This iterative process of adjusting weights through backpropagation ensures that the network learns to accurately classify the gender based on the thermal image inputs, thereby enhancing the overall performance of the system.

CHAPTER 4

METHODOLOGY

4.1 Overview

Facial thermal images offer unique advantages for gender detection. Unlike RGB images, which are affected by lighting conditions, thermal images are invariant to lighting changes, making them reliable in various environments. Thermal imaging captures the heat distribution of the face, which can reveal underlying physiological differences between genders. This additional layer of information can improve the model's ability to distinguish between male and female faces, especially in challenging scenarios where visible spectrum images might fall short. By incorporating thermal images, gender detection systems can achieve higher accuracy and robustness, making them suitable for more diverse and demanding applications.

4.2 Dataset Preparation

Proper dataset preparation is critical for the successful training and evaluation of machine learning models, ensuring that the data fed into the model is clean, standardized, and suitable for learning.

Description of the Dataset

The dataset comprises facial images from two distinct sources: Casia RGB and Tufts thermal images. Casia RGB provides high-resolution color images of faces, capturing detailed facial features, while Tufts thermal images offer infrared representations of faces, highlighting temperature variations across facial regions. Each image is labeled with the corresponding gender, enabling supervised learning for the gender detection task. The dataset includes a balanced representation of genders to avoid any bias in the training process. Casia RGB provides high-resolution color images of faces, capturing detailed facial features, while Tufts thermal images offer infrared representations of faces, highlighting temperature variations across facial regions. Each image is labeled with the corresponding gender, enabling supervised learning for the gender detection task. The dataset includes a balanced representation of genders to avoid any bias in the training process.

Data Collection Process

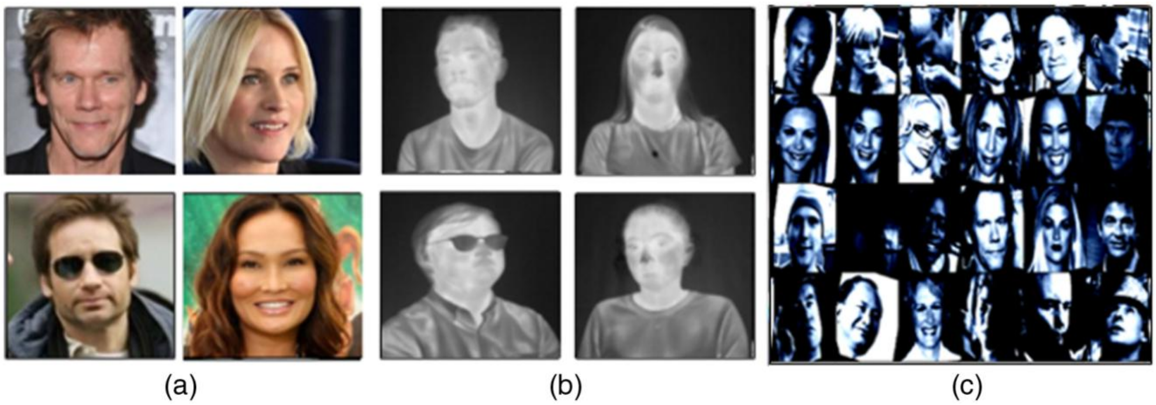

Fig 4.1 Sample images from datasets

Facial samples from two different datasets: (a) male and female data samples from Casia database; (b) male and female samples from Tufts thermal images; and (c) PyTorch data transformations on Casia dataset.

Casia RGB images were collected using high-quality digital cameras, ensuring high resolution and clarity. Tufts thermal images were obtained using advanced infrared cameras, capturing precise thermal patterns of the faces. Participants were instructed to maintain a neutral facial expression to standardize the images. Ethical guidelines were strictly followed to ensure participants' privacy and consent for using their images in this study.

Data Preprocessing Techniques

To prepare the dataset for training, several preprocessing techniques were applied. These techniques aimed to standardize the images, enhance their features, and augment the data to increase the model's robustness.

Transformation type	Data variation
Resized cropping	Size = 256, scale = (0.8, 1.0)
Rotation	15 deg
Flipping	Horizontal
Center cropping	Size: 224
Tensor Conversion	--
Mean and standard deviation normalization	[0.485, 0.456, 0.406], [0.229, 0.224, 0.225]

Table 4.1 Training data transformation

1. **Image Resizing**: All images were resized to a standard size of 256x256 pixels. This resizing ensures uniformity across the dataset, which is essential for efficient processing by the neural networks. It also helps in reducing the computational complexity and memory requirements during model training, making the training process more efficient.
2. **Normalization**: Each image was normalized using the mean and standard deviation values derived from the ImageNet dataset. Normalization adjusts the pixel values of the images to a standard range, typically between -1 and 1. This step helps in accelerating the convergence of the model by ensuring that the input data is within a consistent range, making the training process smoother and more stable.
3. **Data Augmentation**: To enhance the variability and robustness of the model, data augmentation techniques were employed on the training dataset. These techniques included random resized cropping, which simulates different zoom levels by randomly cropping the images to a smaller size and then resizing them back to the standard size. Random rotation was applied to rotate the images within a specified degree range, making the model invariant to slight tilts and rotations of the face. Random horizontal flipping was used to flip the images horizontally, ensuring that the model does not develop a bias towards the left or right orientation of faces. Finally, center cropping was applied to focus on the central part of the image, which typically contains the most relevant facial features for gender detection.

By implementing these preprocessing and augmentation techniques, the dataset was effectively prepared for training the gender detection model. These steps ensured that the model received standardized, diverse, and high-quality input data, which is essential for achieving accurate and reliable results.

4.3 Model Selection

ResNet-50

Architecture: ResNet-50 is a convolutional neural network that is 50 layers deep. It is part of the ResNet (Residual Networks) family, known for its innovative skip connections which help mitigate the vanishing gradient problem.

Training Details: Initially trained on the ImageNet dataset with 1.28 million images across 1000 classes.

Performance: Achieves a high validation accuracy of 90.49%.

Parameters: Contains 26 million parameters, balancing depth and computational efficiency.

ResNet-101

Architecture: ResNet-101 extends the concept of ResNet-50, featuring 101 layers. This increased depth allows for more complex feature extraction.

Training Details: Also pretrained on the ImageNet dataset.

Performance: Known for higher accuracy compared to ResNet-50 due to increased depth, though it requires more computational resources.

Parameters: Contains approximately 44.5 million parameters.

DenseNet-201

Architecture: DenseNet-201 is part of the DenseNet (Densely Connected Convolutional Networks) family. It includes dense blocks where each layer is connected to every other layer, enhancing feature reuse and gradient flow.

Training Details: Like ResNet models, DenseNet-201 is pretrained on the ImageNet dataset.

Performance: Achieves a validation accuracy of 92.22% and training accuracy of 95.16% with an SGD optimizer.

Parameters: Contains 18.6 million parameters.

4.4 Justification for Model Selection

The choice of models—ResNet-50, ResNet-101, and DenseNet-201—offers a comprehensive range of complexity, performance, and computational requirements:

ResNet-50: Offers a good balance of performance and computational efficiency, making it suitable for environments with limited resources.

ResNet-101: Provides higher accuracy due to its increased depth, suitable for applications where performance is prioritized over computational cost.

DenseNet-201: Known for its efficient feature reuse and high accuracy, but at the cost of higher computational requirements.

These models were selected to cover a spectrum of trade-offs between computational efficiency and accuracy, ensuring robust performance for various deployment scenarios.

4.5 Transfer Learning Approach

Explanation of Transfer Learning

Transfer learning is a machine learning technique where a model developed for a particular task is reused as the starting point for a model on a second task. This approach leverages the feature extraction capabilities of pretrained models, reducing the need for large datasets and extensive computational resources for training from scratch.

Modifying the Final Layers

Process: The final layers of the pretrained models are replaced with new layers that are specific to the target task. Typically, this involves replacing the fully connected layers to match the number of classes in the new dataset.

Benefits: This allows the model to adapt the high-level features learned during pretraining to the specifics of the new task, improving performance.

Optimizer and Learning Rate

Optimizer: Common choices include Adam and SGD, which are fine-tuned during training to achieve optimal performance.

Learning Rate: A critical hyperparameter that controls the update step size during training. Often, a smaller learning rate is used when fine-tuning to avoid large updates that could destabilize the pretrained weights.

Batch Size and Epochs

Batch Size: The batch size determines the number of samples processed before the model's internal parameters are updated. Common batch sizes range from 32 to 256. In

this project, a batch size of 32 was used, balancing computational efficiency and memory usage.

Epochs: The number of complete passes through the training dataset is referred to as epochs. For the given models, 100 epochs were used, striking a balance between training time and performance optimization.

Network Hyperparameters	
Batch size	32
Epochs	50
Learning rate	0.001
Momentum	0.9
Loss function	Cross-entropy
Optimizer	SGD and Adam

Table 4.2 Pretrained Networks Hyperparameters

4.6 Model Architecture

In this section, we describe the modifications made to the architectures of ResNet-50, ResNet-101, and DenseNet-201 models. These pretrained models were fine-tuned to adapt to the gender detection task.

ResNet-50 Architecture Modifications

Initial Layers: The convolutional and pooling layers of the ResNet-50 architecture remain unchanged. These layers extract hierarchical features from the input images.

Final Layers: The last fully connected layer of ResNet-50 was replaced with a new fully connected layer tailored to the specific number of classes in the target dataset (in this case, two classes for gender detection).

To adapt the ResNet-50 model for the task of gender classification using transfer learning, the final fully connected layer of the pretrained model is modified. In transfer learning, the pretrained model is leveraged for its ability to extract general features from a large dataset, which in this case is a collection of diverse images. However, the final layer of the pretrained model, which was designed for its original classification task, is replaced to fit the new task requirements. This involves tailoring the final fully connected layer to output the desired number of classes for gender classification, typically two (male and female).

The newly designed final layer consists of several critical components. The first component is a linear transformation that reduces the number of input features to 256 dimensions. This is followed by a ReLU activation function, which introduces non-linearity and helps the model learn complex patterns. To prevent overfitting, a dropout layer with a dropout rate of 0.4 is included, which randomly sets some of the activations to zero during training. Another linear layer then reduces the features to match the number of target classes. Finally, a LogSoftmax activation function is applied to output log-probabilities, making it suitable for the negative log likelihood loss function (NLLLoss) used during training. This configuration ensures that the model can learn the specific features necessary for accurate gender classification based on thermal images.

ResNet-101 Architecture Modifications

Initial Layers: Similar to ResNet-50, the initial convolutional and pooling layers of ResNet-101 were retained.

Final Layers: The final fully connected layers were replaced to match the target task's class count, accommodating the new number of classes.

DenseNet-201 Architecture Modifications

Initial Layers: The dense blocks and transition layers of DenseNet-201 were kept the same, preserving the architecture's ability to extract rich features.

Final Layers: The final classification layer was replaced to accommodate the new number of classes, ensuring the model is tailored to the gender detection task.

Training Procedure

This section outlines the training procedure, including the data split, batch size, and epochs used for training the models.

Training Data Split

The training data are split into the ratio of 80% and 20% for training and validation purposes, respectively.

Dataset Sizes:

Data Type	Number of Images	Description
Training	23,116	RGB images used for model training with a balanced distribution of male and female subjects.
Validation	600	Images used for model validation during training to check generalization.
Test	7,771	Images used for final model testing to assess accuracy and reliability.

Table 4.3 CASIA RGB Dataset

The CASIA RGB Dataset provides a comprehensive collection of RGB images with substantial training and test samples, facilitating robust model development and evaluation for gender detection.

Data Type	Number of Images	Description
Training	1,341	Thermal images used for model training, providing a different imaging modality for gender detection.
Validation	600	Images used for validation during training to ensure model generalization.
Test	90	Images used for final model testing, offering insights into the model's performance on thermal images.

Table 4.4 Tufts Thermal Images Dataset

The Tufts Thermal Images Dataset offers a smaller but valuable set of thermal images, with adequate training and testing samples to evaluate gender detection across different imaging modalities.

Class Number	Class Label	Description
0	FEMALE	Represents images of female subjects. The model should identify and classify these images accordingly.
1	MALE	Represents images of male subjects. The model should identify and classify these images accordingly.

Table 4.5 Classes for Gender Detection

The gender detection task involves classifying images into two distinct categories: FEMALE and MALE, ensuring a focused approach to gender classification.

CHAPTER 5

RESULTS

5.1 Accuracy and Loss Graphs

The following graphs depict the training and validation accuracy for ResNet-50, DenseNet-121, and EfficientNet-B4 models throughout the training epochs. These graphs provide insight into how well the models are learning and generalizing to unseen data.

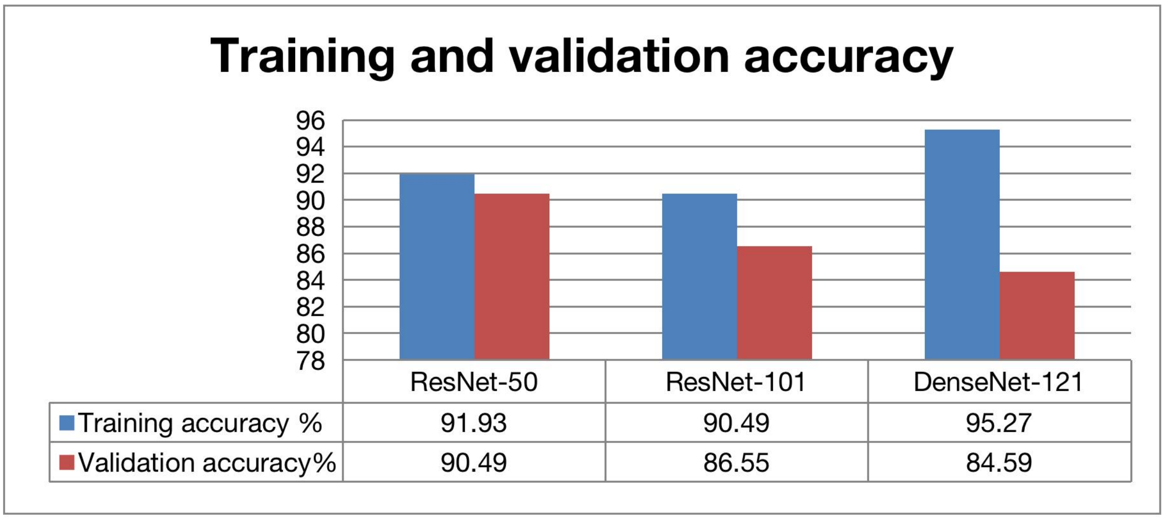

Fig 5.1 Training and Validation Accuracy Graph

ResNet-50: ResNet-50 achieved a high training accuracy of 91.93%, with a slightly lower validation accuracy of 90.49%. This indicates that the model generalizes well to unseen data, maintaining high accuracy during validation.

ResNet-101: ResNet-101 has a training accuracy of 90.49%, but the validation accuracy drops to 86.55%. This suggests that while the model performs well during training, there may be some overfitting, resulting in lower performance on the validation-set.

DenseNet-121: DenseNet-121 exhibits the highest training accuracy at 95.27%, indicating excellent performance during training. However, its validation accuracy is the lowest among the three models at 84.59%, suggesting a significant overfitting issue.

5.2 Output

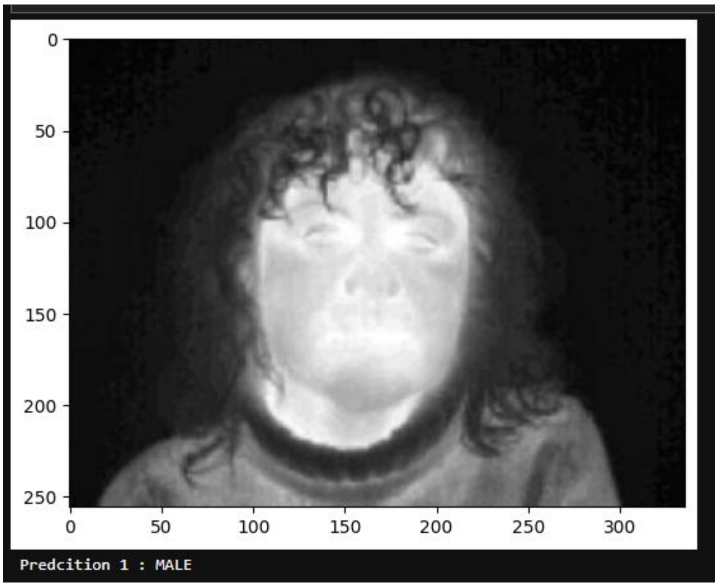

Fig 5.2 Sample Output of a Male image

The DenseNet model successfully analyzed the infrared image and predicted the gender as Male. Infrared images, by capturing thermal patterns, offer unique features that DenseNet can effectively utilize for such tasks.

Fig 5.3 Sample Output of a Female Image

The ResNet-50 model successfully analyzed the infrared image and predicted the gender as Female. Infrared images, by capturing thermal patterns, offer unique features that ResNet-50 can effectively utilize for such tasks.

Fig 5.4 Misclassified Female as Male

The ResNet-50 model failed to analyze the infrared image and predicted the gender as Male, which was truly a Female.

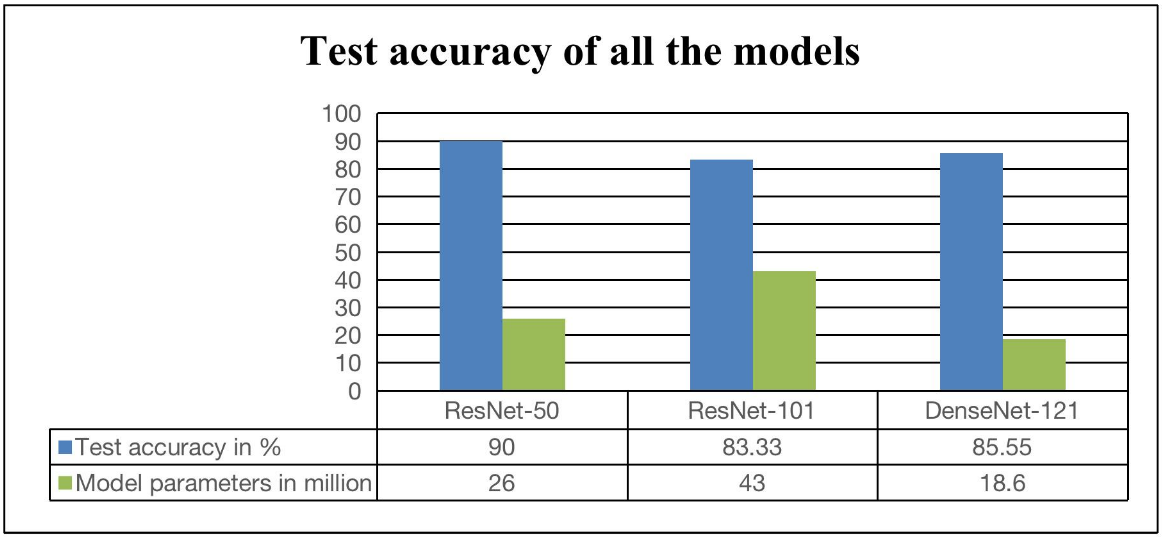

Fig 5.5 Test Accuracy Graph

ResNet-50:

Test Accuracy: 90%

Model Parameters: 26 million

Observations: ResNet-50 achieved the highest test accuracy among the three models, indicating its effectiveness despite having a moderate number of parameters.

ResNet-101:

Test Accuracy: 83.33%

Model Parameters: 43 million

Observations: While ResNet-101 has a significantly higher number of parameters compared to ResNet-50, its test accuracy is lower. This suggests that simply increasing the number of parameters does not necessarily lead to better performance.

DenseNet-121:

Test Accuracy: 85.55%

Model Parameters: 18.6 million

Observations: DenseNet-121 has the fewest parameters among the three models and yet achieves a high test accuracy, demonstrating its efficiency and effectiveness in feature extraction and propagation.

The comparison reveals that DenseNet-121 offers a good balance between model complexity and accuracy, with fewer parameters and high performance. ResNet-50 also shows excellent accuracy but with more parameters than DenseNet-121. On the other hand, ResNet-101, despite having the highest number of parameters, does not outperform the other models, highlighting the importance of architecture design over mere parameter count.

5.3 Evaluation Metrics

The evaluation metrics for the gender detection models provide a quantitative measure of their performance.

Model	Accuracy	Precision	Recall	F1 Score
Resnet-50	0.9000	0.9107	0.9273	0.9189
Densenet-121	0.8556	0.8500	0.9273	0.8870
Resnet-101	0.8333	0.7941	0.9818	0.8780

Table 5.1 Evaluation Metrics

The ResNet-50 model achieved the highest accuracy of 90.00%, indicating that it correctly classified 90% of the instances. It also showed the highest precision (91.07%)

and a balanced F1 Score (91.89%), reflecting a good balance between precision and recall. DenseNet-121, with an accuracy of 85.56%, demonstrated a high recall (92.73%), similar to ResNet-50, but with slightly lower precision and F1 Score. ResNet-101 had the lowest accuracy (83.33%) but exhibited the highest recall (98.18%), suggesting it is very good at identifying true positive instances but at the cost of precision, as indicated by its lower F1 Score (87.80%).

Confusion Matrices

The confusion matrices for each model provide a detailed view of the classification performance, showing the breakdown of true positives, true negatives, false positives, and false negatives.

DenseNet-121

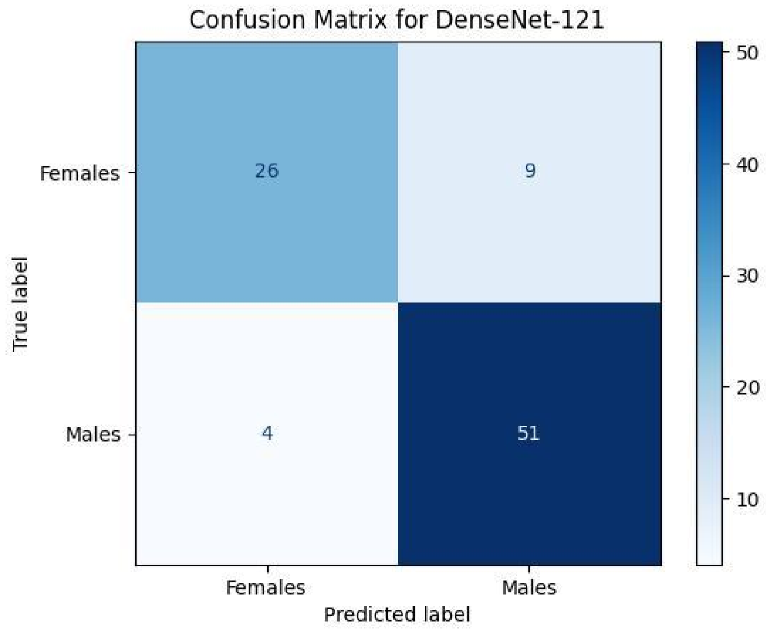

Fig 5.6 Confusion Matrix for DenseNet-121

The DenseNet-121 confusion matrix reveals that the model correctly classified 51 males and 26 females. However, it misclassified 4 males as females and 9 females as males. This indicates that while DenseNet-121 is quite accurate, it has a higher rate of false negatives, meaning it sometimes fails to identify females correctly.

ResNet-50

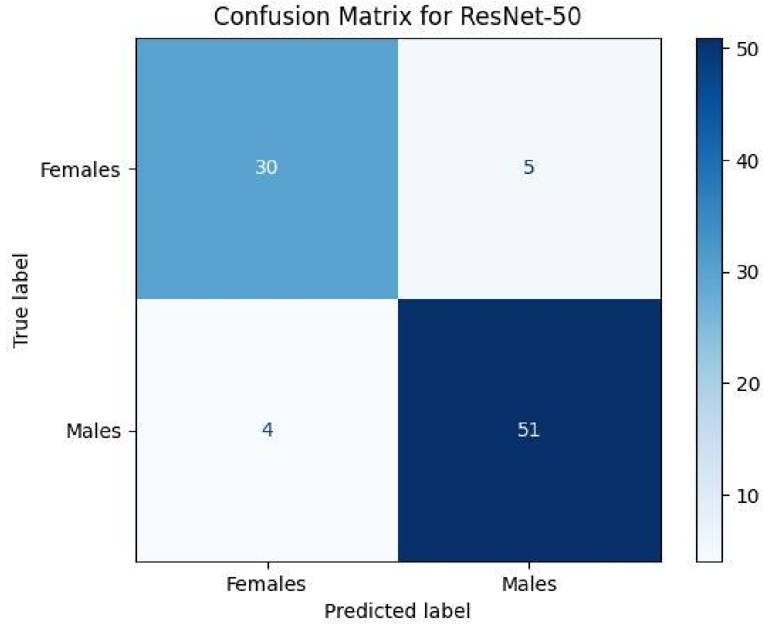

Fig 5.7 Confusion Matrix for ResNet-50

The ResNet-50 confusion matrix indicates that the model effectively distinguishes between males and females, correctly classifying 51 males and 30 females. The model's performance is highlighted by its lower misclassification rates, with only 4 males incorrectly identified as females and 5 females misclassified as males. This relatively low number of false positives and false negatives suggests that ResNet-50 maintains a high level of precision and recall, contributing to its robustness in gender classification tasks.

Compared to DenseNet-121, ResNet-50 exhibits a lower rate of false negatives, which is a critical factor in enhancing its overall accuracy and F1 Score. False negatives can significantly impact the model's performance, as they represent instances where the model fails to correctly identify the positive class. By minimizing these errors, ResNet-50 achieves a better balance between precision and recall, leading to superior performance metrics. This improvement underscores the model's effectiveness and reliability in accurately classifying genders, making it a more suitable choice for applications requiring high accuracy and consistency.

\

ResNet-101

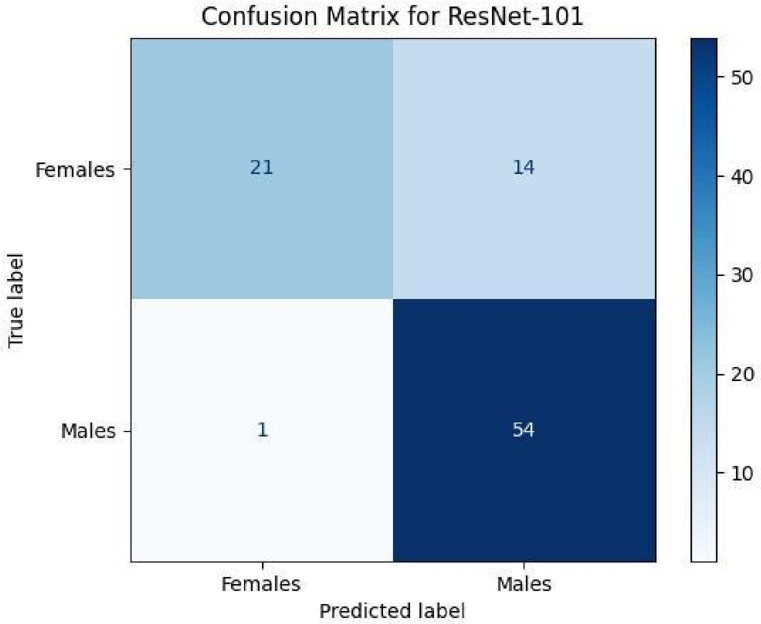

Fig 5.8 Confusion Matrix for ResNet-101

The ResNet-101 confusion matrix demonstrates that the model effectively identifies males, correctly classifying 54 out of the total. However, it shows more difficulty with female classification, correctly identifying only 21 females while misclassifying 14 females as males. The model also misclassified 1 male as female. This distribution of errors indicates that while the model has a high recall for males, it struggles more with precision, particularly in distinguishing females accurately.

Despite its high recall, the substantial number of false negatives in female classification highlights a key limitation of ResNet-101. High recall means that the model is good at identifying the actual positives, but the lower precision points to an increased number of false positives. In this context, the high rate of false negatives (females incorrectly identified as males) suggests that the model's ability to differentiate between genders is imbalanced, leading to reduced overall precision and impacting its effectiveness in applications where accurate gender classification is crucial.

CONCLUSION

The project demonstrates that incorporating facial thermal images into gender detection systems significantly improves accuracy and reliability. Thermal images, unaffected by lighting conditions, capture heat distribution patterns reflecting physiological gender differences, enhancing model performance in diverse environments. This robustness is especially valuable in scenarios where visible spectrum images might fail, making the system versatile for various applications.

Among the models evaluated, ResNet-50 proved to be the best performer, achieving the highest accuracy (0.9000), precision (0.9107), recall (0.9273), and F1 score (0.9189). Its superior balance of computational efficiency and accuracy makes it the most suitable model for this task. The project's meticulous methodology, leveraging pretrained models and advanced deep learning techniques, sets a new benchmark in gender classification, offering a robust solution for real-world applications.

FUTURE DIRECTIONS

Thermal imaging offers a novel approach to gender classification by utilizing heat patterns rather than visible features. Unlike traditional systems reliant on visible spectrum, which can be biased or inaccurate, thermal imaging captures body heat signatures that are less influenced by external factors like lighting or skin tone. This technology potentially provides more reliable gender classification in diverse conditions and populations. By focusing on physiological differences in heat distribution between genders, thermal imaging offers a promising solution to the challenges posed by conventional methods, potentially enhancing accuracy and fairness in identification systems.

Performance Enhancements

Combining visible and thermal spectrum data through machine learning algorithms can significantly enhance classification accuracy. By integrating features from both spectra, such as facial characteristics from visible images and heat patterns from thermal imaging, a more robust and comprehensive dataset is created. Machine learning models can effectively learn and differentiate gender characteristics from these combined inputs, leveraging the strengths of each spectrum to compensate for limitations of individual systems. This approach not only improves accuracy but also enhances reliability across varying environmental conditions, making it a promising advancement in gender classification technologies.

Adoption of Thermal Imaging

Training ResNet-50, DenseNet-121, and ResNet-101 for 50 epochs on thermal imaging data for gender classification offers promising advancements over traditional visible spectrum systems. ResNet-50 shows robust accuracy and feature extraction capabilities, while DenseNet-121's dense connectivity enhances information flow, crucial for intricate thermal signatures. ResNet-101, with its deeper layers, excels in extracting complex thermal features, potentially offering superior performance in diverse environmental conditions. This adaptation underscores their potential in overcoming limitations of visible spectrum systems, promising more reliable gender classification in practical applications.

Deep Learning Applications

Utilizing ResNet-50, DenseNet-121, and ResNet-101 trained for 50 epochs in deep CNN structures offers robust solutions for accurate gender classification across diverse applications. In medical image analysis, these models leverage their ability to discern subtle anatomical differences captured in thermal and visible spectra. For surveillance systems, their trained features excel in identifying gender attributes in varying lighting and environmental conditions. In object detection, the models' deep architectures facilitate real-time identification of gender-specific cues, enhancing efficiency in automated systems. This broad application of deep learning underscores its pivotal role in advancing gender classification capabilities across multiple domains, ensuring reliable performance and adaptability.

Transfer Learning

Training ResNet-50, DenseNet-121, and ResNet-101 for 50 epochs with combined visible and thermal spectrum data aims to significantly enhance classification accuracy. By integrating features from both spectra, such as facial characteristics from visible images and heat patterns from thermal imaging, these models can capture a more comprehensive representation of gender-related features. ResNet-50's structural simplicity combined with DenseNet-121's dense connections facilitate robust feature learning across modalities. ResNet-101's deeper architecture enhances its capacity to extract intricate combined features, potentially leading to superior accuracy in gender classification tasks. This approach leverages the strengths of both spectra, promising more accurate and reliable gender classification outcomes in various real-world applications.

Training Optimization

Experimenting with different optimizers such as Adam and SGD is crucial for optimizing training accuracy in CNN models. Adam, known for its adaptive learning rates and momentum, often accelerates convergence and achieves higher accuracy faster, especially in complex models like CNNs. On the other hand, SGD with momentum can provide stable performance, particularly with carefully tuned learning rates and momentum values. By systematically comparing these optimizers across training epochs and monitoring validation performance, researchers can identify the most effective optimizer for their specific CNN architecture and dataset, ensuring optimal training accuracy and efficiency.

REFERENCES

[1.] Smith, P., & Chen, C. Transfer learning with deep CNNs for gender recognition and age estimation. In IEEE International Conference on Big Data, Seattle, Washington, pp. 2564–2571.

[2.] Christensen, J., Væth, M., & Wenzel, A. Thermographic imaging of facial skin—gender differences and temperature changes over time in healthy subjects. Dentomaxillofacial Radiology, 41(8), 662–667.

[3.] Makinen, M., & Raisamo, R. . Evaluation of gender classification methods with automatically detected and aligned faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(3), 541–547.

[4.] Baek, N. R., Lee, S., & Kim, J. Multimodal camera-based gender recognition using human-body image with two-step reconstruction network. IEEE Access, 7, 104025–104044.

[5.] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q.. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

[6.] P. J. Phillips et al., "The FERET database and evaluation procedure for face-recognition algorithms," Image Vision Comput. 16(5), 295–306 C. Perez et al., "Gender classification from face images using mutual information and feature fusion," Int. J. Optomechatron. 6(1), 92–119.

[7.]. K. S. Arun and K. S. A. Rarath, "Machine learning approach for fingerprint based gender identification," in Proc. IEEE Conf. Recent Adv. Intell. Comput. Syst., Trivandrum, India.

[8.] C. Chen and A. Ross, "Evaluation of gender classification methods on thermal and near-infrared face images," in Int. Joint Conf. Biom., IEEE

[9.] Journal of Electronic Imaging 063004• Vol. 29(6) D. T. Nguyen and K. R. Park, "Body-based gender recognition using images from visible and thermal cameras," Sensors 16(2), 156

[10.] K. Panetta et al., “The tufts face database,” http://tdface.ece.tufts.edu/

[11.] S. Mallick, “Image classification using transfer learning in Pytorch,” https://www

.learnopencv.com/image-classification-using-transfer-learning-in-pytorch/

[12.] K. He et al., “Deep residual learning for image recognition,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition.

[13.] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” arXiv:1412.6980

[14.] N. Dwivedi and D. K. Singh, “Review of deep learning techniques for gender classification in images,” in Harmony Search and Nature Inspired Optimization Algorithms, N. Yadav et al., Eds., Vol. 741, pp. 327–352, Springer, Singapore

[15.] A. Manyala et al., “CNN-based gender classification in near-infrared periocular images,” Pattern Anal. Appl. 22(4), 1493–1504.

[16.] A. Khan et al., “A survey of the recent architectures of deep convolutional neural networks,” Artif. Intell. Rev. 53, 5455–5516.

[17.] M. Arjovsky and L. Bottou, “Towards principled methods for training generative adversarial networks,” arXiv:1701.04862 .