Multi-Task Learning with Soft Parameter Sharing Using MobileNetV2 for Body Type and Skin Tone Classification for Recommend Clothing.

MSc Research Project

MSc Artificial Intelligence

Nayana Dorai Raj

Student ID: 23418991

School of Computing

National College of Ireland

Supervisor: Ade Fajemisin

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Nayana Dorai Raj		
Student ID:	23418991		
Programme:	MSC Artificial Intelligence	**Year:**	2025
Module:	Practicum		
Supervisor:	Ade Fajemisin		
Submission Due Date:	11.12.2025		
Project Title:	Multi-Task Learning with Soft Parameter Sharing Using MobileNetV2 for Body Type and Skin Tone Classification for Recommend Clothing.		
Word Count:	………………………………………	**Page Count**	………………………………………………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Nayana Dorai Raj
Date:	10.12.2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

NCI

Practicum – 2

Report

Multi-Task Learning with Soft Parameter Sharing Using MobileNetV2 for Body Type and Skin Tone Classification for Recommend Clothing.

Nayana Dorai Raj

23418991

Abstract

The proposed paper presents body-type and skin tone classification from a female full-body image with dual-phase computer vision system that combines unsupervised representation learning and supervised multi-task classification with soft sharing parameter to allow automated fashion recommendation that is tailored to each user. In the initial phase there is an unsupervised pipeline that carries out body-type discovery and skin-tone clustering based on the use of feature extraction, PCA dimensionality reduction and K-Means clustering with a refinement step made through K-Nearest Neighbour smoothing. Estimating skin-tone uses the luminance-invariant YCrCb chrominance statistics, region-of-interest sampling, and body-type clustering uses pose-normalised geometric descriptors to cluster shapes to form a consistent phenotype group, with no labelled supervision. On the second stage the labels of unsupervised learning is passed to the next step, a multi-task learning (MTL) neural architecture simultaneously predicts a body type and dermatologically appropriate skin-tone classes through late fusion of body spatial features and facial skin patches, sharing soft-parameters to optimise cross-task generalisation. Results of learning streams are combined into a silhouette-colour recommendation system by which morphological and chromatic features are associated with proven styling guidelines to women fashion. Experimental findings suggest that involving unsupervised clustering will give accurate results by avoiding manual labelling to achieve representation coherence, and supervised MTL to achieve discriminative accuracy, results in stronger phenotype estimation, fairness across under-represented tones, and consistency in automated outfit generation than individually run single-task and supervised baseline methods do. The accuracy obtained for skin tone and body type are 62.96% and 81.48% respectively with Multi-task learning.

Key words: **Unsupervised Learning, Supervised Learning, KNN, K-Means, MobileNetV2, Multi-task Learning with Soft Parameter Sharing, PCA dimensionality reduction, YCrCb chrominance, RGB, Sampling, Balanced sampling, Joint-balanced sampling, Fitrizpatrick Classification, Unsplash API, Gradio interface, Deep Learning, and CNN.**

1. INTRODUCTION

Fashion is an intensely subjective and highly expressive sphere, but over the decades fashion industry has been highly dependent on generalised sizing, fixed rules on the colour usage and the garment design guided by the trends. On the one hand, online commerce allows bringing clothes to any part of the world that suits, however, on the other hand, online shopping presents new customers with problems: they often do not realize how clothes will fit their bodies, what a color will suit them, or whether clothes will fit into their own style. According to surveys of leading retailers, over 35 percent of returns online are always related to poor fit or dissatisfaction with appearance, which is one of the indications that there is still a gap between the expectations of consumers and the available recommendation mechanisms.

Machine Learning, Deep Learning and artificial intelligence has the potential to fill this gap. Over the past few years, the application of artificial intelligence (AI) to fashion technology has grown at an accelerated rate, with the growing need to provide personalization in digital experiences in e-commerce, virtual try-on functions, and consumer styling apps. Convolutional neural networks (CNNs), representation learning, and multi-task design (S Ruder et al.(2017)) have allowed visual attributes such as body proportions, color harmony, facial features and the styling context to be analyzed on a single photograph. Nevertheless, the majority of the current tech fashion solutions use manual annotation, face recognition alone, or demographic data that cannot reflect the complexity of personal styling. On the same note, commercial tools tend to suggest clothing only to statistical correlations learned, and not considering the established fashion principles in the sense of silhouette, proportion, and chromatic unity (CL Istook et.al (2001)). The motivation behind this study is that there is a gaping deficiency in the literature: it is lacking a fully automated, end-to-end, personalized fashion recommendation system which can operate based on body-type analysis and skin-tone estimation. Moreover, an effective dataset of labelled body types and skin tones can be frequently missing, which leads to the necessity of a new data generation pipeline that is capable of automatically labelling unstructured images with a hybrid feature extraction and clustering strategy.

This project develops a whole, AI-based fashion recommendation system to solve these shortcomings. It uses pose estimation along with silhouette width measurements, deep visual encodings, statistical skin-tone modelling, color-expert learning branches, sharing of soft parameters, MobileNetV2(M Sandler et al. (2018)), and multi-task training to offer personalized clothing recommendations. The system provides both deployable and holistic prototype of virtual styling assistants through the combination of real-world fashion rules by using curated style advice and a live image search based on the Unsplash API.

Research Question: How accurately can a recommendation system be developed using image input to analyse both body type and skin tone to suggest personalized clothing styles for female users, using deep learning models with multi-task learning?

2. Literature Review

Recently, there has been advances in the field of fashion recommendation systems (FRSs) to further advance beyond basic collaborative filtering technique to methods that integrate features of clothes, physiological data, and visual appearance. This section critically analyses some of the major contributions of researchers in the domain of personalized fashion recommendation, body-shape and skin-tone based analysis and the extraction of visual features. All the literature reviewed works are evaluated in relation to their goals, shortcomings and their applicability to the proposed model, the combination of the body type and skin-tone based recommendation using multi-task learning and MobileNet architecture.

2.1 DeepFashion Dataset

Liu et al. (2016) introduced the DeepFashion dataset, a foundational asset for fashion recognition, featuring more than 800,000 labeled images cataloging clothing categories, and attributes. This large scale annotated dataset facilitates the training of models that thrive on large, finely annotated input. Its extensive scope and depth have driven much of the recent advancement in the fashion AI field. For the present study, DeepFashion's wide-ranging, densely annotated content remains invaluable, furnishing a solid foundation for the development of models that demand substantial labeled fashion data. Nonetheless, the dataset omits annotations for skin tone and body shape, restricting its immediate utility for the specific aims of the current research.

Recently, a review of the current methods of fashion recommenders was done by Deldjoo et al. (2023) which shed light on the weaknesses of most models which only rely on user-item interactions overlooked the visual and contextual aspects of the user such as body shape or skin tone among others.

2.2 Personalized Clothing based on Body Shape and Clothing Style

Hidayati et al. (2018) introduced a system where the fashion style suggestions use body shape recognition by extracting body proportions in the images and correlating them to style categories. Their value is their explicit modeling of body features as well as the proposal of types of clothes (like dresses or trousers) that suits body type. The model built classifies the user user's body type by analyzing full-body image. This task is performed using Convolutional Neural Network architecture. This involves image processing for body type analysis, deep feature extraction for clothing items, and supervised learning to suggest outfits.

In comparison, Pandit et al. (2020) reviewed clothing recommendation systems that work on user attributes and found out that most previous systems to consider were general heuristics or surveys as opposed to computer vision-based body shape recognition. This comparison makes the approach used by Hidayati one of the rare examples of using a vision as a part of a personalized solution, but it also remains devoid of other powerful features such as skin tone or the context of a certain event. This review paper emphasizes the implementation of decision tree classifiers, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) in attribute based recommendations. Some authors have implemented clustering algorithms to group similar users or styles.

The above studies demonstrate the effectiveness of body shape based recommendations.

2.3 Skin Tone and Context Aware Fashion Recommendation

Garude et al. (2019) presented a system, which suggests the outfits by skin tone and occasion. Their approach is based on mapping color palette verses skin tone categories and occasion tagging to match outfits. The significance of this contribution lies in the fact that it openly rallies around the relevance of outfit selection based on cultural and contextual background. The implementation uses computer vision for skin tone detection and a decision based logic framework for generating outfit suggestions.

Nevertheless, Chakraborty et al. (2021) also conducted a review of fashion recommendation approaches, and in their work, they highlighted that such systems as those of Garude are not integrated or deep learning presented and real-time use of user images, but rather that they are based on pre-determined inputs of users. This constraint impacts on scalability and user experience. The authors have highlighted the role of deep learning midels in Fashion Recommender systems. CNNs are used for feature extraction from the images to enable visual similarity based recommendations. Recurrent Neural Network (RNNs) and LSTMs help capture sequential behavior of user and fashion trends. Autoencoders are used for dimensionality reduction and latent feature learning, Generative Adversarial Networks (GANs) are applied to generate new fashion suggestions and complete the outfit suggestion.

Tangwei Ye et al. introduced a scene-conscious fashion recommender system that makes recommendations about fashion depending on a scenario (e.g., party, gym). The authors have used new dataset called Scene-Aware Fashion Recommendation Dataset (SAFR), which has 85,000 outfit-scene pairs which are manually labeled with compatibility. The model incorporates environmental factors in order to become more realistic. Although effective where the situation is concerned, it fails to consider personal biometric information. ResNet CNN is used to extract visual features from scene images, and CNN to process outfit images. A multi-model module is implemented to integrate both scene and fashion features. The outfits are recommended on the compatibility scores.

In comparison with Jing et al., which focuses on the issue of visual-category relations, Tangwei Ye et al. put an accent on external context. The approaches are complementary, and neither of them focuses on internal user-specific characteristics. This comparison helps provide evidence towards our goal of implementing the concept of the biometric inputs into recommended fashions. For this experiment the authors have used Polyvore and Polyvore-D datasets. A pre-trained CNN extracts image features. Transformer-based encoder is used for item descriptions. Each category is embedded into a semantic vector. A multimodal us used to align text features and image for each item.

The above papers considers skin tone and occasion.

2.4 Monocular 3D Pose and Body Shape Estimation

Zanfir et al. (2018) pointed out a model of 3D pose and body shape estimation based on monocular images in natural scenes based on multi-scene constraints. This method provides better estimates of physical characteristics, and this is important in getting body data of the user out of standard RGB images.

However, Liu et al. (2016), although in the first attempt to perform a visual clothing retrieval, failed to model the 3D human body features and poses, which represents a technical gap. Combining the estimation process of Zanfir and fashion recommendation might increase specificity of body-friendly fashion recommendations. The authors survey includes the Deep learning models such as CNN, R-CNN, Siamese-CNN, and RNN implementation by various researchers. This survey highlights the limitations of most models that depend only on either skin tone or body type. Additionally, hybrid models combine collaborative filtering with contentment-based filtering, to benefit user interaction history. CNNs and RNNs are integrated to focus on salient features and refine recommendation outputs. GNN are adopted to model brand associations, and style compatibility.

Hence, the body shape estimation and monocular systems should be therefore combined with clothing recognition and recommendation systems to personalize fashion recommendation to its full extent.

2.5 Surveying Modern Fashion Recommendation Systems

Both Deldjoo et al. and Chakraborty et al. (2021) conducted a review of recent inventions in the area of fashion recommender systems. Deldjoo et al. also focused on content-based filtering with visual deep learning characteristics. A comparative study of collaboration filtering, hybrid and computer vision-based was given by Chakraborty et al. An important lesson of the two reviews is that multimodal signals (visual, textual, contextual) are important to successful personalization.

When comparing the two works, Deldjoo is relatively more recent and technically all-inclusive whereas the review of Chakraborty lays emphasis on the level of categorization at the application level. These findings contribute to the inclusion of the visual and contextual characteristics, such as the shape of the body and the skin color, into a single recommendation pipeline in the context of this study. Combined these reviews enhance the validity of our research direction: merging profound visual analysis of body shape and skin tone with situational suggestions of personal fashion recommendations.

De Divitiis et al. created an aspect disentanglement model that breaks down the clothing attributes and makes them independent factors that easily determine the recommendations along the independent dimensions. Their model enables the recommendation process to be transparent, which is ignored in the black-box deep learning systems. A variational Autoencoder (VAE) framework is used with factor specific latent representation.

The old literature has highlighted rich techniques to recommend an outfit based on the body-shape and skin-tone and occasion and to estimate pose through images, as well as pointing out the drawbacks of the existing recommenders. However, since some of the readily available datasets such as DeepFashion are well annotated when it comes to clothes, they do not provide body type and skin tone related attributes.

2.6 Fritzpatrick based Skin Tone classification

The recent developments in AI-based dermatology have highlighted the issue of models which can be trusted to work under the entire range of skin tones. The framework suggested by Ulrich et al. (2025) represents a step forward in quantifying skin-tone, replacing the dichotomous Fitzpatrick scales by transforming melanin-related chromatic signatures by decomposing YCrCb colour-space, averaging over space, and preprocessing the results in ways that are illumination-robust. They combine convolutional neural networks with calibrated regression heads to generate continuous pigmentation scores, with which more granular dermatological risk stratification can be done. This pipeline also includes histogram normalization and colour constancy algorithms to reduce lighting bias, and this is able to do the cross-tone generalization much better.

Aggarwal et al.(2021) performed a comparative analysis of the most popular CNN models, such as ResNet-50, Inception-v3, and DenseNet-121, on more difficult Fitzpatrick classes (IV-VI). The researchers found out a significant deterioration of performance by class imbalance, variability of reflectance and reduced contrast at melanin-rich skin. Aggarwal measured the effect that dataset skew has on sensitivity and had revealed that models that are trained on lighter skin color in bulk fail to capture texture features and lesion edges in darker ones. The results imply the need of training sets with balanced tones, adversarial illumination enhancement, and colour-neutral feature encoders. Light-skinned trained models were always more accurate at detecting melanoma and BCC, and had higher sensitivity, specificity, F1 scores, and AUC, which indicates that there are performance disparities and lower diagnostic accuracy on darker skin-of-color images.

These works combined highlight the role of pigmentation-conscious preprocessing, continuous tone modeling, and bias-reduction methods in the development of a fair dermatological AI diagnostic system.

2.7 Sampling Techniques

Hordri et al. (2018) introduce a methodological review of methods of resampling to solve the extreme class imbalance problem in credit card fraud detection in each case, in which cases of fraudulent transactions constitute less than 1 percent of the overall samples. Some of the resampling paradigms compared in the study are Random Undersampling (RUS), Random Oversampling (ROS), and Synthetic Minority Oversampling Technique (SMOTE), as well as hybrids, such as SMOTE with Tomek Links, and SMOTE with Edited Nearest Neighbours (ENN). The authors build a supervised classification pipeline technically, where resampled datasets are inputted into machine learning classifiers, most of which are decision-tree-based models and logistic regression. SMOTE works by creating synthetic minority samples by interpolation in feature space with the Tomek Links and ENN steps eliminating borderline or noisy majority samples to increase the separability of classes. It trains the models with 10-fold cross-validation and evaluates their results based on imbalance-sensitive measures, such as Precision, Recall, F1-score and Area Under the ROC Curve (AUC). Findings indicate that hybrid resampling (particularly SMOTE + ENN) can significantly enhance the detection of fraud because it lowers the number of overlapping areas and enhances the representation of minority classes without significantly increasing noise. The paper concludes that with cautious use of synthetic oversampling with noise-reduction techniques, more robust and generalisable classifiers are available on highly skewed datasets.

2.8 MobileNetV2

The paper presents a lean convolutional neural network (CNN) framework, which is mobile-based and based on the classification of artificial skin-types on the face, to classify the skin type using MobileNetV2 framework, to be deployed on smartphones. The inverted residual blocks and depthwise separable convolutions of MobileNetV2 make it highly cost-effective in terms of computational power (they can be run in real-time on devices with low power consumption), and highly accurate (the model can be utilized in a variety of applications). The pipeline consists of face detection, region-of-interest (ROI) extraction, and preprocessing, i.e. illumination normalization and resizing to align with the input resolution of MobileNetV2. The model is conditioned using a filtered set of labeled facial images of typical dermatological skin types, with data augmentation performed to make the model resistant to lighting and pose changes. Transfer learning is used by pre-training MobileNetV2 using ImageNet weights and training the last layers on dermatological classification. The trained model is then exported in the form of an efficient execution in Android devices using the TensorFlow Lite. It was experimentally demonstrated that the MobileNetV2-based system provides high-quality, low-latency skin-type classification that can be applied to a realistic definition of mobile dermatology.

3. Methodology

3.1 Dataset

The data is a set of 821 unlabeled full-body images, sourced through two publications (300 images of DeepFashion-Multimodal dataset, Kaggle and the rest of 521 images, Pinterest). The DeepFashion subset offers better images in terms of their frontal pose and the same background whereas the images in Pinterest present a more diversified background in terms of poses, lighting, clothing fashion, and real-life situation. There were no body-type and skin-tone labels, so the dataset is completely unlabeled, unsupervised learning is performed to generate labels for both skin-tone and body-type.

3.2 Unsupervised Learning.

The unsupervised part of this study seeks to find natural structure in the data, which will be performing clustering on the individuals in terms of visual characteristics of body shape and skin pigmentation. This step is conducted before the stage of supervised learning, and it is not based on any ground-truth labels. The objective is to determine natural clusters which can be related to patterns of the body-types and also patterns of skin tones.

3.2.1 Data Pre-processing

- Every image is subjected to a standardised pre-processing pipeline to denoise and normalise visual inputs:
- Consistency in fixed spatial dimensions Image resizing to smaller spatial dimensions across the dataset.
- Detection of face and body regions on the basis of pre-trained OpenCV/MediaPipe models to detect pertinent areas.
- Alternation and equalization of pixel values to stabilize feature extraction.
- Conversion of colour space RGB to YCrCb, allowing chrominance-based analysis of the variation in melanin and pigmentation.

3.2.2 Feature Extraction

Visual data that could be significant in body shape and skin coloration is obtained by extracting two feature groups:

a. Body Shape Features

- Shape-ratio (width-to-hip ratio, shoulder line curvature)
- Proportions of the bounding boxes (height/width ratios)
- Histogram of Oriented Gradients (HOGs) Silhouette descriptors.
- PCA dimensionality reduction to keep only major structural features optional.
- Such features encode geometric clues that describe categories of shapes like pear, rectangle or upside down triangle.

b. Skin Tone Features

Cr/Cb thresholds in YCrCb space are used to identify regions of skin. For each detected region:

- Cr and Cb mean values are obtained.
- An approximation has to be made on the intensity of melanin using the chrominance distribution.
- Colour histograms are reduced to low dimensional vectors.
- This makes it possible to represent Fair/Medium/Dark in a continuous feature space instead of subjective categorical labels.

3.2.3 Clustering Approach

K-Means is used separately on:

- Body-type clusters
- Skin-tone clusters
- The selection of number of clusters is done in an empirical manner as follows:
- Elbow method (minimizing with-in cluster sum of squares)

- Silhouette (cluster compactness and separability) analysis.

The algorithm organizes images in categories of similarity of geometrical and chromatical characteristics. Each image is assigned:

C body = f (shape features)

C skin =f (chrominance features)

These labels of clusters are subsequently converted into pseudo-labels to aid supervised learning or dataset balancing.

3.2.4 Post-processing

Cluster results undergo:

- Filtering to combine small or thin clusters.
- Map centroid attributes to meaningful attributes (e.g. warm and cool undertones, hourglass like figure).

The unsupervised pipeline therefore generates groups of data that can be interpreted by humans without having to have labelled data.

3.3 Multi-Task Learning (MTL) with soft parameter sharing.

Multi-task learning model is a collaborative learning architecture that learns to classify body type and skin tone shared using a common deep neural architecture. This method harnesses shared lower-level visual representations (e.g. shape contours, colour representations independent of lighting) to enhance both learning speed and accuracy.

Multi-task learning will be justified by the following reasons:

- The two classification tasks rely on lower-level common visual semantics.
- Body-type recognition is aided by edge structures and silhouette shapes.
- The estimation of skin-tone relies on effective colour representation and uniform correction of illumination.
- The shared feature extractor minimises redundancy and enhances generalisation and the task specific heads concentrate on their own goals.

3.3.1 Pre-Processing

- The unsupervised learning pipeline of pre-processing is likewise employed.
- Image normalization and resizing.
- YCrCb colour conversion
- Body/face region extraction

- Data augmentation (rotation, zoom, brightness jitter) to avoid overfitting, which occurs randomly.

3.3.2 Shared Feature Extraction Backbone

The goal of shared feature extraction backbone is to determine features that are shared by both sets of elements or objects in question. All images are sent to a CNN-based backbone (MobileNetV2) to come up with a shared representation:

- The hierarchical features are obtained through convolution layers.
- Normalization of batches makes training stable.
- ReLU activation maintains non-linearity.
- Global average pooling causes dimensionality reduction.

This common encoder offers a common feature vector:

$$F_{shared} = E_{\theta}(X)$$

3.3.3 Two Task Specific output heads.

The similar feature set is inputted into two parallel streams:

a. Body type classification Head:

- Dense layers
- Dropout for regularization
- The output of the softmax with n classes of body-types.

$$\hat{y}_{body} = f_{body}(F_{shared})$$

b. Skin Tone Classification Head:

- Dense layers
- Wild colour sensitive regularization.
- Softmax results of Type I- Type VI tone (or more levels by your labels).

$$\hat{y}_{skin} = f_{skin}(F_{shared})$$

3.3.4 Multi-Task Loss Function

The two activities are optimised together with weighted loss:

$$L_{total} = \alpha L_{body} + \beta L_{skin}$$

where:

L_{body} =body type categorical cross-entropy loss.

L_{skin} = skin tone categorical cross-entropy loss.

α and β control the significance of each task.

These weights have been adjusted to the optimum value.

3.3.5 Training Procedure

- The end-to-end training of the model is through Adam optimiser.
- Termination of early monitors loss of validation.
- Convergence is stabilised by decay of the learning-rate.

Evaluation measures: accuracy, precision, recall, F1-score of each task. The common backbone acquires general visual representations, and each head is specialised, which is more accurate than training two models in isolation.

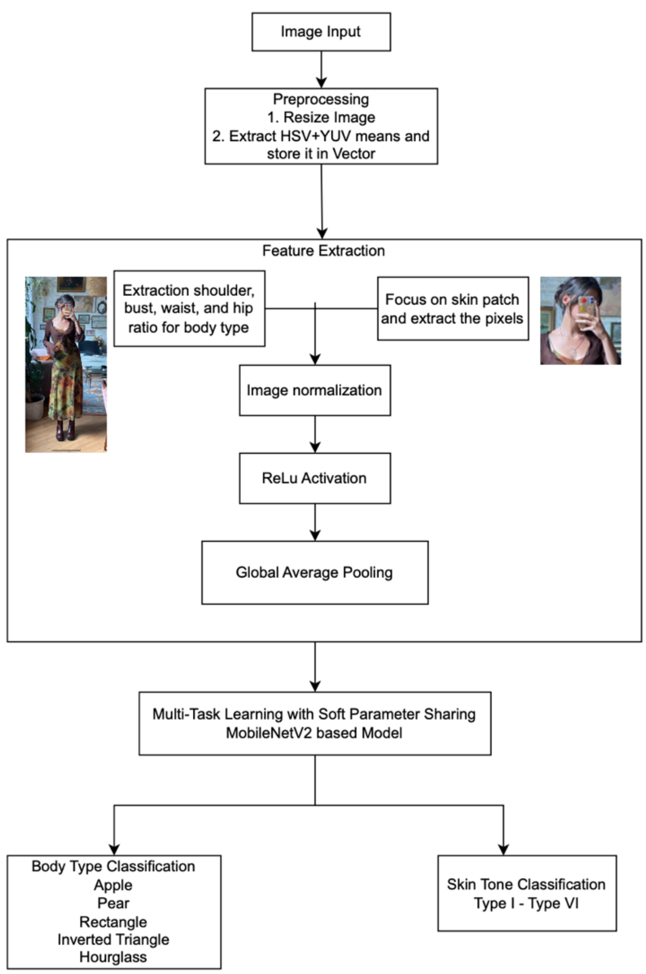

Fig 1: Multi-task Model Architecture

3.4 Outfit Recommendation

Pre-processing is done on the input image including normalization of RGB and resizing. The extraction stream of skin-patches is dedicated to isolate the information of facial chrominance. The facial landmarks are found with MediaPipe FaceMesh, when not found, a fallback crop is obtained. The skin patch and full body image are inputted into a pre-trained multi-task model based on MobileNetV2 which makes simultaneous predictions of body type, and skin tone. It is a multi-task method that facilitates learning of the same low-level features and permits task-specific heads of classification.

The skin tones are mapped to Fitzpatrick types II- VI and each of them formats a bespoke palette of colours that are complementary to the observed chrominance elements. The desired body size is projected onto silhouette-optimised category of outfits (e.g., A-line, wrap, V-neck). A query system with controlled queries is used to recall outfits from Unsplash API, and is built based on fashion-specific search prompts and filters based on semantic text screening and HSV-based colour verification. This will make sure that outfits retrieved will be matching the suggested colour palette, as well as, body-type styling advice.

4 Implementation

4.1 Body Type Classification using K-means clustering:

The technique implemented for classifying the dataset is a hybrid of conventional computer vision, deep learning based feature extraction, anatomical landmark based geometry, unsupervised learning and further refinement using K-nearest neighbour classification. After loading the dataset, the feature extraction for body type classification begins with Mediapipe models, Pose estimation and Selfie segmentation. Pose estimation determines anatomical landmarks, that is, shoulder and hip points, that serve as reference points to calculate the body measurements like bust and waist. Selfie segmentation results in creating a binary representation of the human mask that separates the subject silhouette and the background. The silhouette allows the system to compute four fundamental anthropometric measurements which include shoulder width, bust width, waist width and hip width. These measurements are estimates of actual proportions of human bodies and provide definite geometrical data that have been used in classifying body shapes. Simultaneously, the MobileNetV2 lightweight and yet highly expressive convolutional neural network, which has been trained on ImageNet, is used to process the same image. The model is applied as a feature extractor and only the convolutional base is used. The last average-pooled embedding yields a 1280-dimensional representation of high-level features like shape contours, shading patterns, clothing textures and edges, and general body appearance. These deep features serve to supplement silhouette-based geometric features with more abstract and global cues.

The hybrid feature is the feature vector of every image that is generated by combining the four silhouette measurements with the deep features generated by MobileNetV2. This gives an image a large 1284-dimensional vector. Because high-dimensional representations may have redundant or noisy information the second step is dimensionality reduction by Principal Component Analysis (PCA). The features are then standardized before PCA so that all dimensions make equal contribution. The feature space is reduced by PCA, with 95 percent of the total variance. This measure has a tremendous impact on computational efficiency, stabilization of the clustering process, and preservation of the necessary structural information stored in geometrical and deep-learning parts. The obtained reduced PCA characteristics are the foundation of the body type clustering, which is done through KMeans with five clusters. The five most popular female body types, Apple, Pear, Hourglass, Inverted Triangle and Rectangle, are the number of the clusters. Notably, KMeans is used in the unsupervised mode. The PCA and scaling transforms are reversed to obtain the original-scale centres of each cluster, only the silhouette measurements are obtained based on these centroids since they are the most comprehensible measurements regarding body shape. Each cluster centroid, therefore, represents the average shoulder, bust, waist, and hip width of the group represented by centroid.

Apple

Pear

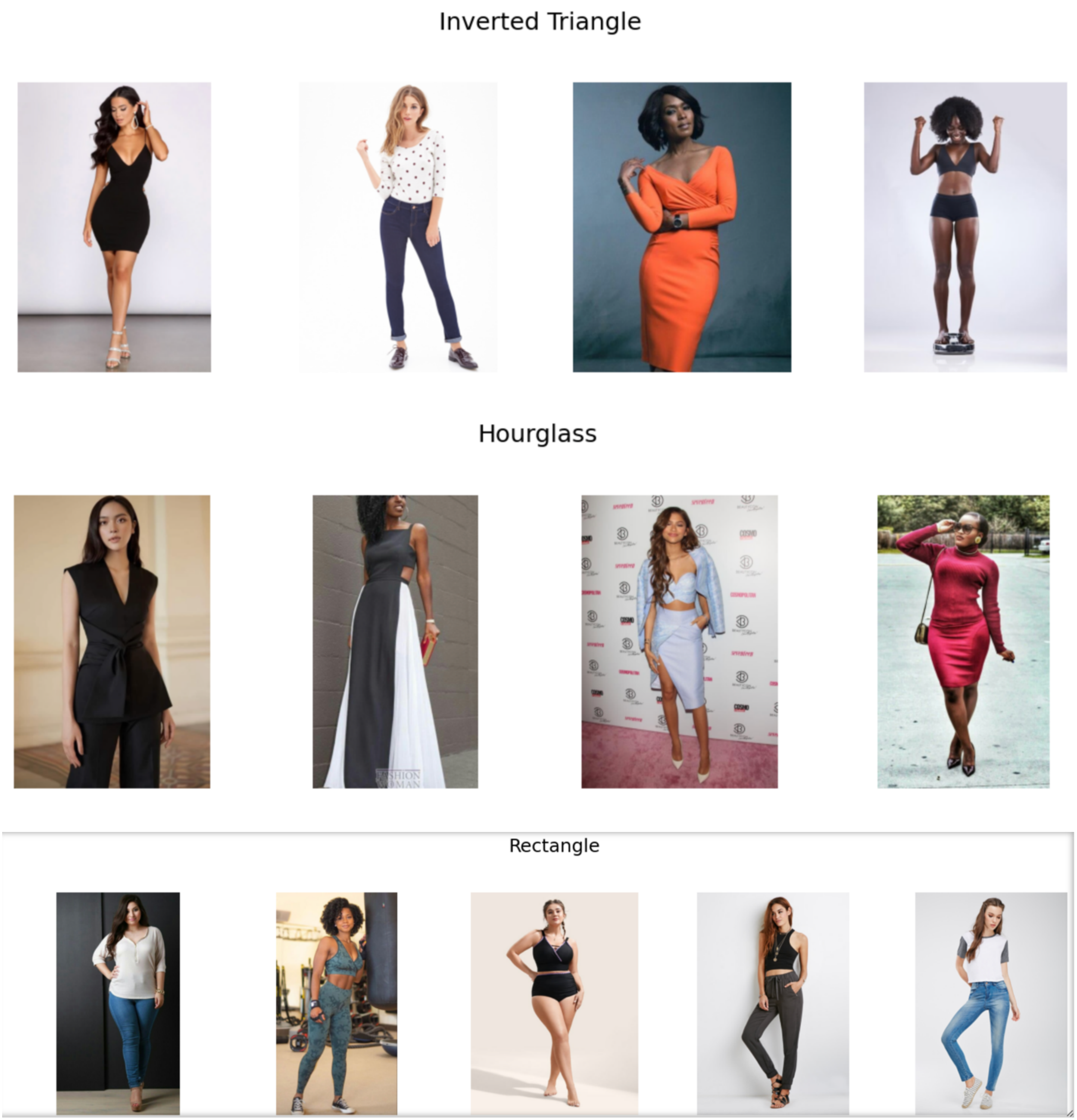

Fig 2: Sample images of Body Type classification

The anthropometric logic as a foundation of the methodology is to map each KMeans cluster to a particular body type using an entirely automatic scoring system. A few of the scoring functions measure how similar each cluster is to the body characteristics. For example, the Apple shapes have been linked with a relatively larger waist measures, the Pear shape with much broader hips, the Inverted Triangle shape with broader shoulders, the Hourglass shapes with evenly measured shoulders and hips and narrow waist, and the Rectangle counterparts with less variation in all measures. Body types are ranked each in their respective clusters using the heuristic scores. A one-to-one assignment algorithm is then used where the five clusters are uniquely assigned to the five body types. With this method, manual labeling is removed and a data-driven scaling mapping strategy is created.

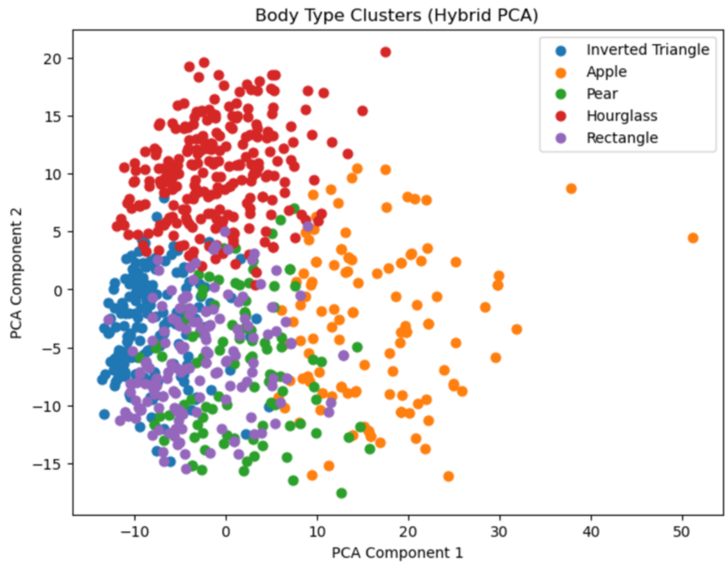

Fig 3: Body Type Classification Clusters

4.2 Skin Tone Classification using KNN

The skin tone classification pipeline starts with the pre-processing of every image and the conversion of the image to a colour space, which is especially effective in the analysis of the melanin variations. The approach does not presuppose the use of the standard RGB form which involves lighting, colour, and saturation but converts every image to the YCrCb colour space. This form of representation divides luminance (Y) and chrominance elements (Cr and Cb), which allows simpler capture of underlying skin pigmentation regardless of the ambient light. The system identifies a central region-of-interest (ROI) in each image to ensure that the system is not affected by other objects in the image such as clothing, hair or backgrounds. The empirical centre is the most probable area to have any exposed skin, particularly the areas of the face, and hence the sampling will be reliable as there is no need to explicitly detect the face. The ROI is then flattened to an array of colour pixels, which is the raw input of the next processing.

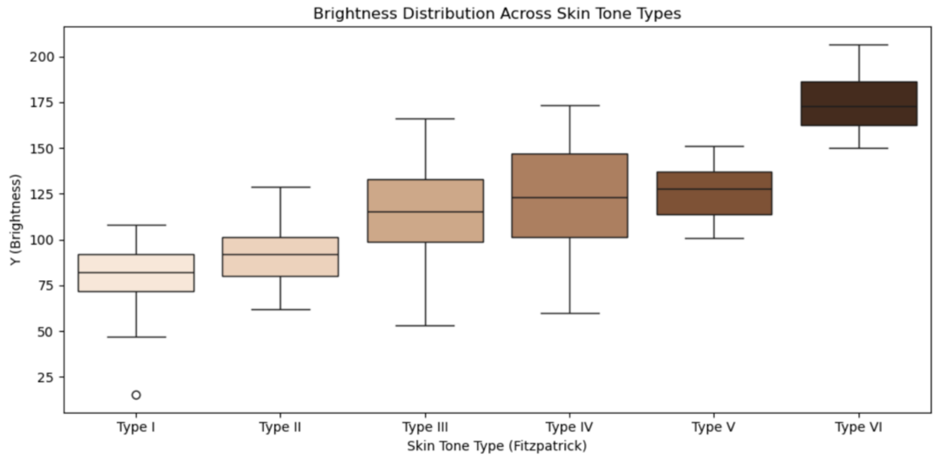

Fig 4: Brightness Distribution

Since the luminance (Y channel) may be highly varied because of the environmental considerations like shadows, brightness or camera exposure, the system offers a filtering mechanism to enhance robustness. No pixels that are too dark or too bright are kept to make sure that classification is based on intrinsic colour and not lighting artefacts. In the case that filtering is excessive, the entire ROI is taken as fallback to avoid data loss. The rest of the pixels are averaged in creating a three-dimensional colour indicator that depicts the average Y, Cr and Cb of the area on the skin of the individual. This small parameter has a biological significance: Cr and Cb represent the relative amount of red and blue chromatic data associated with melanin amount and skin color, and the Y channel preserves relative brightness.

The pipeline standardises the descriptors to prepare them to the clustering process such that every colour dimension has an equal contribution. Standardisation eliminates the variations in scale and normalises distance based operations to make the system interpretable and visualisable, Principal Component Analysis (PCA) is used to compress the number of colour descriptors to two major components. Though not employed in the classification, this provides a visual inspection of the underlying structure of skin-tone variation, which gives a two-dimensional projection of chromatic patterns.

To estimate the six Fitzpatrick categories, the system uses a K-Means clustering model to set up a six data-driven clusters. The K-Means clusters samples according to their similarity in terms of chromatic features of the YCrCb color space, and thus the natural variations in skin tone pigmentation become visible without labelled training. After the clustering process, the groups are analysed based on the mean luminance (Y value). This order is necessary since Fitzpatrick types are classified as a continum between the lightest and the darkest. The clusters are thus ranked in order of brightness with the faintest to the brightest which allows a one-to-one correlation between cluster ID and Fitzpatrick type (I to VI). This automatic ordering makes the classification based on biological progression of phototypes but not arbitrary numbering of clusters. Whilst clustering recovers the pattern of skin-tone variation, a single sample close to a cluster edge may be incorrectly clustered because of noise, or subtle variations in lighting, or unusual chromatic structure. The system also has a K-Nearest Neighbours (KNN) refinement step to enhance the local consistency.

The KNN classifier is taught the cluster configuration found by the K-Means then it reallocates each sample on the agreement of its closest neighbours. This misclassification isolation removal mechanism and cluster cohesion enhances. The fine assignments are further translated into their respective Fitzpatrick classes based on the previous order. The last product connects every processed image with its estimated type of Fitzpatrick skin-tone. Since the system involves the colour-space transformation, unsupervised learning, and neighbourhood refinement, it is also resistant to the changes in lighting, environment and subject matter used in the background. The outcome is a biologically significant, interpretable and completely automated skin-tone classification pipeline, which fits well into the fashion-recommendation system.

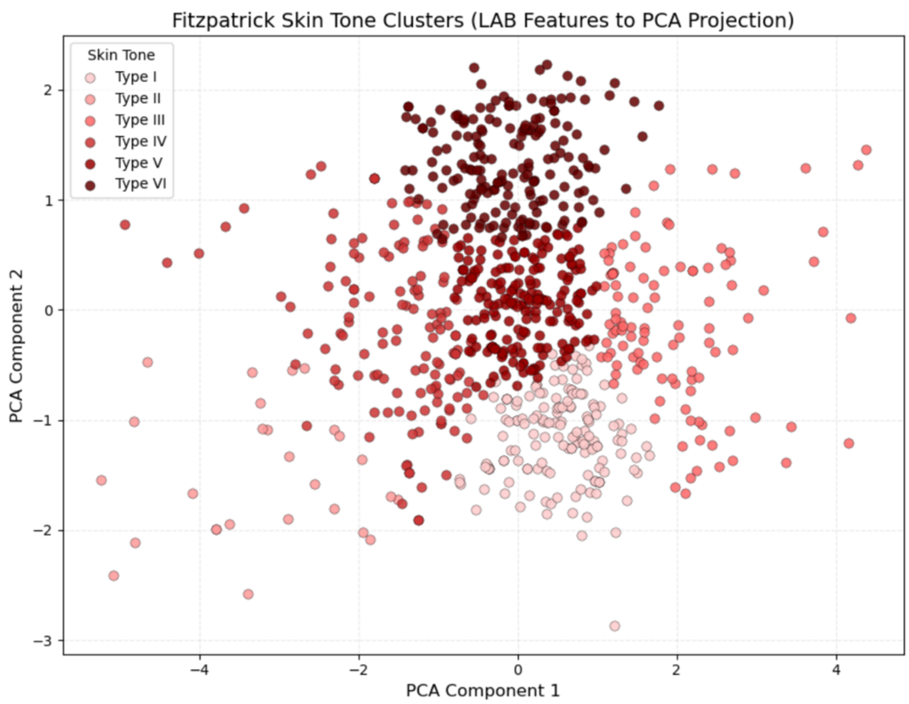

Fig 5: Skin Tone classification clusters.

Body Type	**Counts**
Apple	94
Pear	176
Rectangle	100
Hourglass	126
Inverted Triangle	158

Table 1: Body Type Classification and Counts

Skin Tone	**Counts**
Type I	100
Type II	99
Type III	201
Type IV	186
Type V	49

Type VI	19

Table 2: Skin Tone Classification and counts

4.3 Multi-Task Learning with Soft parameter sharing using MobileNetV2

The algorithm starts with the loading of a dataset and body-type and skin-tone labels obtained from KNN model. Once loaded, the dataset is inspected immediately on the presence of rare classes in the category of body-type. Any of the body types that occur less than three times are dropped. This is a pruning operation that is necessary in deep learning settings where very small classes are unable to give useful gradient information, and they tend to destabilize multi-task learning. After the removal of rare classes, the dataset is filtered, so that only all the images mentioned were valid, uncorrupted and were present at the designated directory. These are only the validated images that are used in training and testing. The labels of body-type and skin-tone are then coded in numeric format with label encoders to allow them to serve as the categorical targets in training. The data is stratified split into a training and a testing partitions where 20 percent is used to test the data. Stratification will provide a guarantee that distribution of body type categories will be consistent across the two sets, which will be necessary to have comparative evaluation particularly where some categories are poorly represented.

Since the two prediction tasks rely on different visual features, the preprocessing pipeline will generate two complimentary views per image. The first is a resized full-body photograph that maintains the entire body proportions with a constant size throughout the data set. This perspective is vital in the acquisition of long-range structural features like the relative width of shoulders, waist and hips that form core in body shape classification. The second one is a patch of skin cut off, which is situated on the upper central part of the shot to the area of the face, neck, and the part that touches the shoulder. This area is selected as it gives a comparatively clear visualization of skin surfaces and it has constant chrominance data without as much interference by clothing or environmental apparatus. To provide a lightweight skin-tone-specific skin patch is resized to a smaller input resolution. Data augmentation is used on both inputs in order to improve the generalization of the model. The augmentation strategy is however well restrained so that it does not distort chrominance values which are important in predicting skin tone. The transformations are therefore restricted to geometric augmentation like random horizontal flips, small rotations and random cropping. Augmentations like the change of brightness, change of hue or saturation are avoided based on intentional exclusion to retain label fidelity. These geometrical extensions enhance the randomness of body postures and backgrounds without losing the integrity of skin tone information.

The model architecture is a soft parameter sharing MTL design. The entire body image is triggered by a convolutional backbone that is pre-trained and it retrieves the high-level global features. Simultaneously, the skin patch is fed through a specialized, low-weight convolutional network that has been trained on skin tone classification color and texture information. The two streams of features are then partially aligned by a soft-sharing mechanism whereby similarity between the latent features spaces of the two tasks is promoted by a regularization term. This method facilitates the exchange of knowledge between tasks as well as each branch having its own task-related specialty. In this respect, soft sharing is especially useful since body morphology and skin characteristics are weakly correlated with each other, and useful in improving learning without committing the model to strict parameter sharing.

The two stages of training ensure that stability is achieved and also overfitting does not occur. During the former stage, the backbone of the feature extraction is fixed and only the task-specific layers and skin-tone subnetwork are trained. This makes sure that the high level visual representations do not change, but the model does learn to match the two tasks. During the second step, fine-tuning of the backbone with selective layers is performed at a lower learning rate that allows the network to tune its high-level features to the combined goals without compromising the advantages of pre-trained initialization. Model performance is measured based on accuracy and detailed classification report, which will include precision, recall, and F1-score per class. This gives a holistic picture of the performance regarding the task specific behavior as well as the efficiency of the multi-task architecture. The trained model is ultimately saved to be used in downstream inference.

The stratification of body-type labels is done so that their distribution is consistent across splits. It first calculates the distribution of the skin tones and then it calculates the distribution of the body types. More to the point, it calculates the combination joint distribution of the skin tone and body type combinations. This joint distribution is paramount since the concept of multi-task learning relies implicitly on the power of the model to comprehend the statistical association between the two jobs. Extremely rare pairs are also a challenge since the model might not see sufficient examples of such combinations in the course of training. Such weak joint cases are strictly determined and subsequent balancing measures are taken to help balance out such imbalances.

Individual images are read, decoded and resized to a 224x224 size that MobileNetV2 needs. A color-constancy operation (gray-world normalization) is performed before the additional processing. In this operation, the average color of a scene in a neutral lighting is supposed to be similar to gray. In the case of a picture with excessive color biases caused by the illumination (e.g. yellow light indoors or blue light outdoors), this normalization will adjust the color channels to bring them to zero. Using this fix makes skin-tone predictions be less reliant on environmental lighting, thus enhancing fairness because of the decrease in illumination-related variance.

To further make the model sensitive to color, the pipeline extracts six handcrafted features of every image, the mean values of both HSV and YUV color spaces. These representations represent luminance, saturation, hue, and chromatic variations and are semantically useful in describing skin tone that is added to the spatially learned patterns of the CNN backbone. All training samples undergo standard augmentations like horizontal flipping and color jittering. Nonetheless, the images of underrepresented classes of skin-tone are more boosted, which imparts variability that enhances generalisation. Moreover, the tones that are extremely rare are augmented with color specifically, as the tones can be undertrained otherwise.

The model is further regularized by mixup augmentation. Mixup uses imagined training samples that are formed as a combination of image pairs and label pairs. The amount of blending is regulated by a mixing coefficient that is generated by a Beta distribution. In the initial stages of training, mixup is used in a forceful manner, which enables the model to be trained in a smoother decision boundary, making it less vulnerable to noise. When training continues the intensities of mixup become zero and the network is able to sharpen its knowledge using more natural data. This annealed mixup approach is a hybrid between the early training during aggressive regularization and of high-fidelity learning during later epochs.

The model architecture is a multi-branch network which is composed of shared and task-specific networks. The most basic component is the MobileNetV2 backbone that is originally trained on ImageNet and it extracts high-level semantic features of the image. The result of its work is transmitted to a common dense layer, which is a basis of the two task heads. The body-type head is comparatively basic and involves the dropout regularization of dense layers, which end with a softmax classifier. The task of this head is to pay much attention to the shape-related cues on a global scale, which MobileNetV2 is inherently good at.

The head of color, on the contrary, is considered much more complicated. The methodology accepts that the proper classification of skin tones cannot be done without a detailed perception of color. The skin-tone branch contains a special color expert subsystem consisting of two streams, one of which is a shallow convolutional network applying local color textures on the input image itself, and another one is a dense transformation of the 6 handcrafted color features. The two streams are joined together to create a rich color properties representation. In order to enable the model to combine shape-related data provided by the shared trunk and color-related data provided by the color expert, multi-head attention module is implemented in the methodology. The color features are projected into a common space and the shared features are projected into a common space and they are considered separate tokens. The attention mechanism also learns cross-interactions among these tokens allowing the network to dynamically compute the extent to which it puts an emphasis on either the shape cues or the color cues in predicting the skin-tone. The other sampling techniques experimented are as follows:

A. Oversampling

Class imbalance, especially in the skin tone labels is a major threat of bias in the training. In order to reduce this, an oversampling method is applied on the training data. In each category of skin tones, the sample size is further increased by repeating the images until all categories become large enough, i.e. equal to the majority category. This oversampling is through repetitions hence the synthesized samples do not have to be unnatural in any manner that guarantees that the underlying chrominance distribution is natural. The balanced dataset is then mixed with the body type dataset by a probability-based sampling mechanism which takes batches that are evenly distributed between skin tone classes. This will make sure that all the categories will be given equal representation in training of the model and that there will be no dominance of prevalent tones.

B. Joint-balanced sampling

The effective problem is to learn the cross-product of the two attributes, rather than balancing body-type classes and skin-tone classes separately. Hence, balancing is done on the joint class distribution. A small dataset is formed based on each combination of skin-tone and body-type. In training, every combination is drawn in such a way that inverses to the frequency of that combination so that rarity combinations get equal representation as common ones. Oversampling is adaptive, too rare combinations are repeated many times, moderately rare combinations are sampled moderately, and common combinations are never repeated. This method will lead to training batches that are evenly distributed representing all the demographic variations without causing biases that would otherwise distort the model.

C. Balanced Sampling

The balanced sampling aimed at removing the statistical imbalances. Body-type and skin-tone prediction within the framework of multi-task learning may bring a lot of skew in the data across different categories, particularly the skin tone. Multiple samples express some tones (e.g., Type III or Type IV), whereas other samples (e.g., Type I and Type VI) can be represented just a few times. Models fit to such distributions are likely to overfit to the majority, making their predictions biased, which do not predict underrepresented populations. Balanced sampling represents the method of providing the network with equal distributions of training signals, which enhances fairness, accuracy, and consistency quality of the network across all demographic segments. Balanced sampling consists of each skin-tone class is represented in every epoch of training in just the same proportion, no matter how many times it occurs in the actual data. The methodology does not randomly sample the images, however, which would give preference to the majority classes, but rather starts by splitting the dataset into individual subsets according to the labels of skin-tone. The skin-tone classes are considered as individual little datasets.

Sampling Method	Skin Tone	Body Type
Joint-Balanced Sampling	0.5122	0.8292
Balanced Sampling	0.5914	**0.8841**
Over Sampling	**0.6296**	0.8148

Table 3: Accuracy of Multi-task model with sampling techniques.

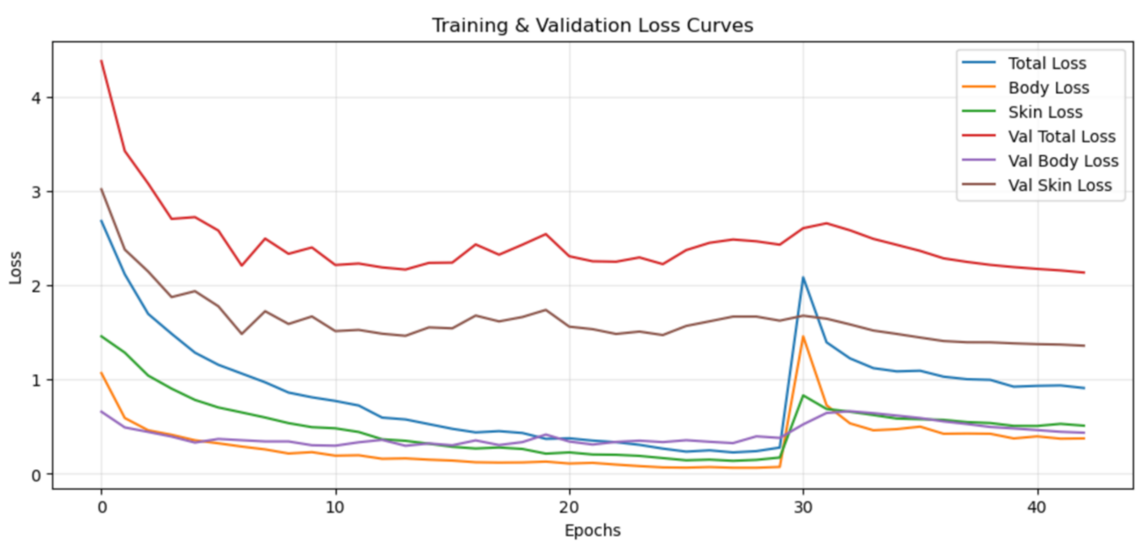

Fig 6: Loss Curves for Body Type and Skin tone during Multi-Task Learning

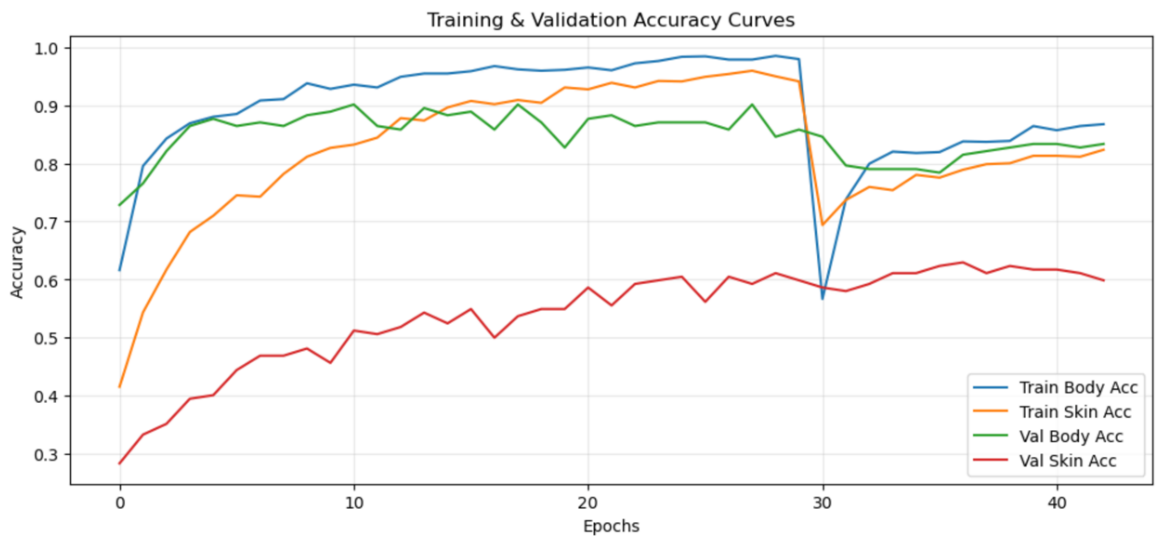

Fig 7: Accuracy Curves for Body Type and Skin tone during Multi-Task Learning

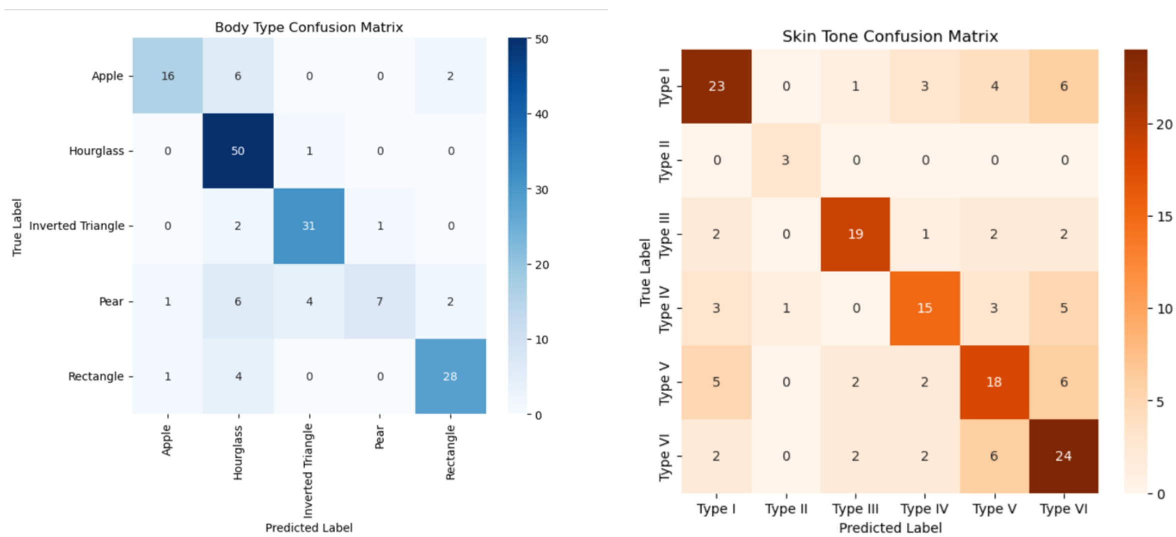

Fig 8: Confusion Matrix for Body Type and Skin Tone

Recommendation Pipeline

The execution incorporates the utilization of TensorFlow, MediaPipe, OpenCV, and Gradio to create an end-to-end smart fashion recommendation platform. A pre-trained multi-task mobile net V2 architecture is loaded into the model, and takes two inputs, a resized full-body image, and an extracted skin patch. Image editing includes conversion to the RGB format, downsizing to the desired input size, and normalization. The skin-patch extraction operates on MediaPipe FaceMesh landmark-based cropping when the facial landmarks cannot be identified, and a fallback central-region crop is applied. The tensors are both inputted to the model to estimate body-type probabilities and skin-tone probabilities. Depending on the results of the argmax, the system will choose between five body types and five Fitzpatrick skin-tone classes. The predictions are mapped onto human understandable descriptors, colour palette and rules of body-shape styling. To retrieve similar outfits, curated query templates are built with the use of predicted style keywords(Fig 8). The system places API requests to Unsplash, retrieves the images, and makes use of stringent semantic filters to make sure that the results are relevant to the fashions of women. Further colour validation on the basis of HSV ensures that the dominant colour tones are the retrieved outfit is supposed to have.

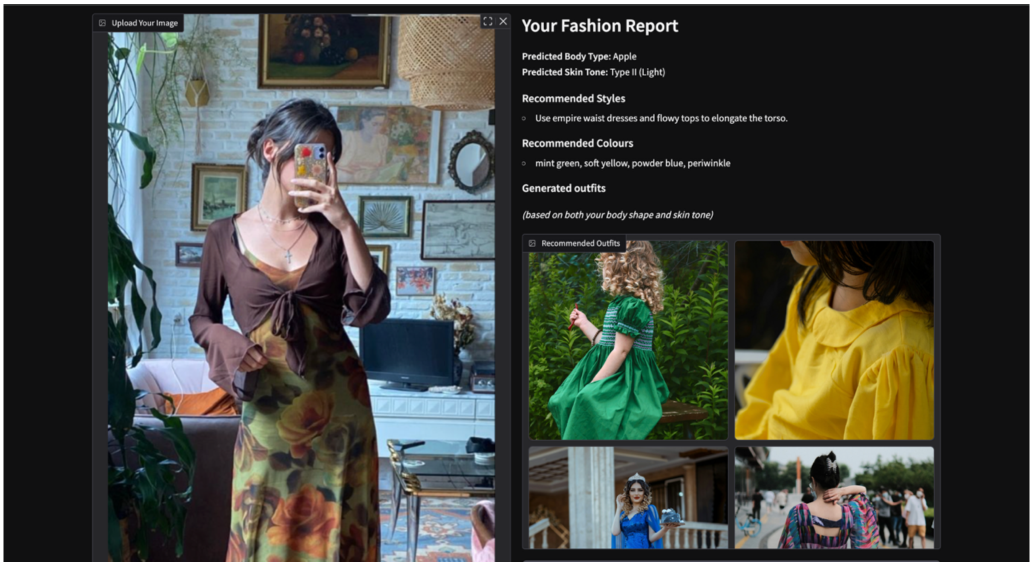

Fig 9: Outfit recommendation based on body-type and skin tone.

5 Conclusion and Future work

The study shows single framework of machine learning that will use unsupervised clustering and the Multi-task learning with soft sharing paraeters (MTL) system based on MobileNetV2 to provide personalised fashion recommendations based on both body morphology and skin pigmentation. The pipeline, which is unsupervised, is successful in revealing latent structure to body-shape geometry and YCrCb-based skin chrominance features to address the difficulty of limited annotations and demographic disproportion. The main contribution however is the multi-task learning model which utilizes a dual-input MobileNetV2 backbone to classify body type and dermatologically compatible skin-tone categories simultaneously. The MTL architecture, based on shared low-level representations, and task-specific dense layers, exploits cross-task generalisation through the encoding of complementary visual information derived by full-body images and localised patches of the skin of the faces, to reduce overfitting and enhance cross-task generalisation. This synergetic sharing of convolutional features allows the model to exceed traditional single-task baselines especially on poorly represented or visually subtle classes. The general system has a high potential in integrating unsupervised representation discovery and deep multi-task inference so that it can produce a high-accuracy recommendation system that balances silhouette-suited styling with a colour theory based on melanin-aware skin tone forecasting. The findings indicate the effectiveness of feature fusion via MobileNetV2 and validate that multi-task learning enhances considerably robustness, fairness, and practical utility of the fashion intelligence systems in the real world. The MTL with joint sampling gave 51.22% and 82.92%, with balanced sampling 59.14% and 88.41%, with image augmentation and oversampling gave 62.96% and 81.48% accuracy for skin-tone and body-type respectively.

For future to add to the accuracy, fairness and practicality of the system, the following changes are to be done:

- Balance the Dataset: Increase the size of the Dataset to balance the class size.
- Enhance Skin Tone Sturdiness: Add illumination-invariant transforms or 3D face modelling or hyperspectral skin modelling to reduce the effects of lighting, shadows, and camera quality further on pigmentation estimation.
- Include Self-Supervised Pretraining: Use contrastive learning (SimCLR, BYOL, MoCo) to improve feature representations in low-label or imbalanced fashion datasets, particularly when using rare body types and wide skin tones.
- Improved Fashion Recommendation Engine: Implement garment segmentation, outfit retrieval and large-scale fashion knowledge graphs, which can be used to of generate richer and context-sensitive recommendation, which can be customized to user tastes, climate and cultural contexts.

References

[1] Istook, C.L. and Hwang, S.J., 2001. 3D body scanning systems with application to the apparel industry. *Journal of Fashion Marketing and Management: An International Journal*, *5*(2), pp.120-132.

[2] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

[3] Ruder, S., 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

[4] Liu, Z., Luo, P., Qiu, S., Wang, X. and Tang, X., 2016. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1096-1104).

[5] Hidayati, S.C., Hsu, C.C., Chang, Y.T., Hua, K.L., Fu, J. and Cheng, W.H., 2018, October. What dress fits me best? Fashion recommendation on the clothing style for personal body shape. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 438-446).

[6] Garude, D., Khopkar, A., Dhake, M., Laghane, S. and Maktum, T., 2019, April. Skin-tone and occasion oriented outfit recommendation system. In *2nd International Conference on Advances in Science & Technology (ICAST)*.

[7] Zanfir, A., Marinoiu, E. and Sminchisescu, C., 2018. Monocular 3d pose and shape estimation of multiple people in natural scenes-the importance of multiple scene constraints. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2148-2157).

[8] Deldjoo, Y., Nazary, F., Ramisa, A., Mcauley, J., Pellegrini, G., Bellogin, A. and Noia, T.D., 2023. A review of modern fashion recommender systems. *ACM Computing Surveys*, *56*(4), pp.1-37.

[9] Pandit, A., Goel, K., Jain, M. and Katre, N., 2020. A review on clothes matching and recommendation systems based on user attributes. *International Journal of Engineering Research & Technology*, *9*(8), pp.786-791.

[10] Chakraborty, S., Hoque, M.S., Rahman Jeem, N., Biswas, M.C., Bardhan, D. and Lobaton, E., 2021, July. Fashion recommendation systems, models and methods: A review. In *Informatics* (Vol. 8, No. 3, p. 49). MDPI.

[11] Jing, P., Cui, K., Guan, W., Nie, L. and Su, Y., 2023. Category-aware multimodal attention network for fashion compatibility modeling. *IEEE Transactions on Multimedia*, *25*, pp.9120-9131.

[12] Ye, T., Hu, L., Zhang, Q., Lai, Z.Y., Naseem, U. and Liu, D.D., 2023, April. Show me the best outfit for a certain scene: A scene-aware fashion recommender system. In *Proceedings of the ACM Web Conference 2023* (pp. 1172-1180).

[13] De Divitiis, L., Becattini, F., Baecchi, C. and Del Bimbo, A., 2023. Disentangling features for fashion recommendation. *ACM Transactions on Multimedia Computing, Communications and Applications*, *19*(1s), pp.1-21.

[14] Delmas, G., de Rezende, R.S., Csurka, G. and Larlus, D., 2022. Artemis: Attention-based retrieval with text-explicit matching and implicit similarity. *arXiv preprint arXiv:2203.08101*.

[15] Chia, P.J., Tagliabue, J., Bianchi, F., Greco, C. and Gonçalves, D., 2022. " Does it come in black?" CLIP-like models are zero-shot recommenders. *arXiv preprint arXiv:2204.02473*.

[16] Wu, Y., Macdonald, C. and Ounis, I., 2022, September. Multi-modal dialog state tracking for interactive fashion recommendation. In *Proceedings of the 16th ACM Conference on Recommender Systems* (pp. 124-133).

[17] Dissanayake, H.P. and Manukalpa, J.C.S., 2025. A Deep Learning Framework for Personalized Fashion Recommendations Based on Skin Tone Analysis. *International Journal of Latest Technology in Engineering, Management & Applied Science*, *14*(6), pp.1012-1019.

[18] Sattar, H., Pons-Moll, G. and Fritz, M., 2019, January. Fashion is taking shape: Understanding clothing preference based on body shape from online sources. In *2019 IEEE winter conference on applications of computer vision (WACV)* (pp. 968-977). IEEE.

[19] Hordri, N.F., Yuhaniz, S.S., Azmi, N.F.M. and Shamsuddin, S.M., 2018. Handling class imbalance in credit card fraud using resampling methods. *Int. J. Adv. Comput. Sci. Appl*, *9*(11), pp.390-396.

[20] Ulrich, P., Zink, A., Biedermann, T. and Sitaru, S., 2025. Beyond Fitzpatrick: automated artificial intelligence-based skin tone analysis in dermatological patients. *npj Digital Medicine*, *8*(1), p.378.

[21] Aggarwal, P., 2021. Performance of artificial intelligence imaging models in detecting dermatological manifestations in higher Fitzpatrick skin color classifications. *JMIR dermatology*, *4*(2), p.e31697.

[22] Hordri, N.F., Yuhaniz, S.S., Azmi, N.F.M. and Shamsuddin, S.M., 2018. Handling class imbalance in credit card fraud using resampling methods. *Int. J. Adv. Comput. Sci. Appl*, *9*(11), pp.390-396.

[23] Saiwaeo, S., Arwatchananukul, S., Mungmai, L., Preedalikit, W. and Aunsri, N., 2023. Human skin type classification using image processing and deep learning approaches. *Heliyon*, *9*(11).

[24] Crawshaw, M., 2020. Multi-task learning with deep neural networks: A survey. *arXiv preprint arXiv:2009.09796*.

[25] Rohani, M., Farsi, H. and Mohamadzadeh, S., 2025. Advanced multi-task learning with lightweight networks and multi-head attention for efficient facial attribute estimation.

[26] Thong, W., Joniak, P. and Xiang, A., 2023. Beyond skin tone: A multidimensional measure of apparent skin color. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4903-4913).

[27] Nurapipah, N. and Yuliana, S.S., 2025. Skin Tone Classification in Digital Images Using CNN For Make-Up and Color Recommendation. *Journal of Intelligent Systems Technology and Informatics*, *1*(3), pp.78-85.