A picture containing text, graphics, font, graphic design

Description automatically generated

Personalized Fashion at a Glance: An Algorithm Tailored to Skin Tone and Body Shape

A thesis submitted in fulfilment of the

requirements for the award of the degree of

MASTER OF SCIENCE IN ARTIFICIAL INTELLIGENCE

|  |  |
| --- | --- |
| **Student Name** | KOH JIA YI |
| **TP Number** | TP072780 |
| **Intake Code** | APDMF2211AI |
| **Programme** | MSc in Artificial Intelligence |
| **Supervisor** | Assoc. Prof. Dr. Imran Medi |

November 2023

Table of Contents

[LIST OF FIGURES 4](#_Toc150439767)

[LIST OF ABBREVIATIONS 6](#_Toc150439768)

[Abstract 7](#_Toc150439769)

[1 Introduction 8](#_Toc150439770)

[1.1 Research Background 9](#_Toc150439771)

[1.2 Problem Statement 10](#_Toc150439772)

[1.3 Research Questions 11](#_Toc150439773)

[1.4 Aim and Objectives 12](#_Toc150439774)

[1.5 Scope 12](#_Toc150439775)

[1.6 Significance of the Research 14](#_Toc150439776)

[1.7 Structure of Thesis 15](#_Toc150439777)

[2 Literature Review 16](#_Toc150439778)

[2.1 Introduction 16](#_Toc150439779)

[2.2 Feature Extraction 17](#_Toc150439780)

[2.2.1 Skin Tone Extraction 17](#_Toc150439781)

[2.2.1.1 Challenges of Skin Tone Extraction 18](#_Toc150439782)

[2.2.1.2 Recent Studies on Skin Tone Extraction 18](#_Toc150439783)

[2.2.2 Pose Estimation 21](#_Toc150439784)

[2.2.2.1 Types of Pose Estimation 21](#_Toc150439785)

[2.2.2.2 Challenges of Pose Estimation 22](#_Toc150439786)

[2.2.2.3 Recent Studies on Pose Estimation and its Applications 22](#_Toc150439787)

[2.3 Recommending System 26](#_Toc150439788)

[2.4 Recommending Algorithms 27](#_Toc150439789)

[2.4.1 Unsupervised Recommending Algorithms 29](#_Toc150439790)

[2.4.1.1 KNN 30](#_Toc150439791)

[2.4.1.2 Autoencoder 32](#_Toc150439792)

[2.5 Fashion Based Recommending System 34](#_Toc150439793)

[2.5.1 User Attribute Based Fashion Recommending System 40](#_Toc150439794)

[2.6 Discussion and Comparison of Related Works 43](#_Toc150439795)

[3 Methodology 46](#_Toc150439796)

[3.1 Workflow of the Project 46](#_Toc150439797)

[3.2 Data Preparation 49](#_Toc150439798)

[3.3 Feature Extraction 50](#_Toc150439799)

[3.3.1 Body Proportion Extraction 50](#_Toc150439800)

[3.3.2 Skin Tone Extraction 52](#_Toc150439801)

[3.4 Algorithm Used for Recommendation 54](#_Toc150439802)

[3.4.1 K-Nearest Neighbors 54](#_Toc150439803)

[3.4.2 Autoencoder 56](#_Toc150439804)

[4 Results and Discussion 60](#_Toc150439805)

[4.1 K-Nearest Neighbours 60](#_Toc150439806)

[4.2 Autoencoder 65](#_Toc150439807)

[4.3 Evaluation of Different Classes of Skin Tone 69](#_Toc150439808)

[4.4 Evaluation Across Genders 72](#_Toc150439809)

[4.5 Evaluation with Daily Life Images 74](#_Toc150439810)

[4.6 Comparison with Related Works 76](#_Toc150439811)

[4.7 Limitation 77](#_Toc150439812)

[5 Conclusion and Future Works 78](#_Toc150439813)

[5.1 Conclusion 78](#_Toc150439814)

[5.2 Future Works 79](#_Toc150439815)

[References 81](#_Toc150439816)

[Appendix A: Literature Review Matrix 85](#_Toc150439817)

# LIST OF FIGURES

[Figure 1. An illustration of the isolation of skin regions (Nada B. Ibrahim, M. Selim, & H. Zayed, 2012). 17](#_Toc150468236)

[Figure 2. An illustration of the difference between 2D pose estimation and 3d pose estimation (Zatolokina, 2023) 21](#_Toc150468237)

[Figure 3. (a) Content-based Recommendation System (b) Collaborative Filtering-based Recommendation System. (YuanZhe, 2022). 27](#_Toc150468238)

[Figure 4. This figure demonstrates the application of the k-Nearest Neighbors (kNN) algorithm for retrieving items based on similarity in a feature space. The red dot symbolizes the query input, for which the system is tasked with finding similar items. The surrounding blue dots depict the various data points within the dataset, each positioned according to its feature values. Those blue dots that are directly connected to the red dot signify the five closest neighbours to the query input—determined by setting K=5K=5—illustrating the data points that bear the smallest distances and, consequently, the greatest similarity to the query. 31](#_Toc150468239)

[Figure 5. An illustration that encapsulates the primary functions of fashion recommendation system. (a) item retrieval, similar items are recommended based on the features of the input image (Samantha Jackson, 2019). (b) suggesting complementary items, complementary items are generated with GAN based on the input image (top) (Sudhir & Mithun , 2019). (c) proposing outfit combinations (Huang, 2021). (d) recommending a capsule wardrobe (Ying & Tao, 2017). 34](#_Toc150468240)

[Figure 6. Overview of the Project's Workflow 48](#_Toc150468241)

[Figure 7. An illustration of the extracted key points used in calculating the body proportions. 51](#_Toc150468242)

[Figure 8. Code snippet of the function “calculate\_propootion”, used in calculating the body proportion. 52](#_Toc150468243)

[Figure 9. This illustration outlines the skin tone extraction method: starting with facial detection, it proceeds by discarding non-skin features, converting the facial region to the HSV color model, and finally calculating the average color to determine the individual's skin tone. 53](#_Toc150468244)

[Figure 10. Code snippet of the function extract\_skin\_tone, employed to obtain the skin tone. 54](#_Toc150468245)

[Figure 11. Code snippet of how KNN is used to find the 5 nearest neighbours. 55](#_Toc150468246)

[Figure 12. This figure provides snippets of code illustrating the definition and training phases of an autoencoder. It highlights the setup of the neural network layers, parameter specification, and the execution of the training routine. 57](#_Toc150468247)

[Figure 13. This figure displays code snippets from the 'get\_recommendations' function, showcasing the retrieval of the top 5 images most similar to the input image within the encoded feature space, as determined by the trained autoencoder model. 59](#_Toc150468248)

[Figure 14. This figure demonstrates the dimensionality reduction from a 6-dimensional feature space to a 2-dimensional plane using t-SNE, where the red point signifies the input data, and the blue points represent the dataset. 61](#_Toc150468249)

[Figure 15. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Euclidean distance metric in the feature space. 61](#_Toc150468250)

[Figure 16. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Manhattan distance metric in the feature space. 63](#_Toc150468251)

[Figure 17. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Minkowski distance metric in the feature space. 64](#_Toc150468252)

[Figure 18. Graph of training and validation loss against epoch throughout the training process of the autoencoder. 65](#_Toc150468253)

[Figure 19. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Euclidean distance metric in the encoed feature space. 67](#_Toc150468254)

[Figure 20. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Manhattan distance metric in the encoed feature space. 68](#_Toc150468255)

[Figure 21. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Mitkowski distance metric in the encoed feature space. 69](#_Toc150468256)

[Figure 22. This figure illustrates the system's recommendations of fashion items that complement dark skin tones, highlighting the algorithm's adaptability to skin tone variability. 71](#_Toc150468257)

[Figure 23. This figure illustrates the system's recommendations of fashion items that complement fair skin tones, highlighting the algorithm's adaptability to skin tone variability. 72](#_Toc150468258)

[Figure 24. This figure illustrates the system's recommendations of fashion items that complement female, highlighting the algorithm's adaptability to gender variability. 73](#_Toc150468259)

[Figure 25. This figure displays the system's capacity to recommend fashion items based on images from daily life, demonstrating the robustness and practical application of the recommendation algorithm. 74](#_Toc150468260)

# LIST OF ABBREVIATIONS

AE Autoencoders

AI Artificial Intelligence

ANN Artificial Neural Network

CNN Convolutional Neural Network

GAN Generative Adversarial Network

HSV Hue, Saturation and Value

KNN K-Nearest Neighbours

RNN Recurrent Neural Network

SVM Support Vector Machine

# Abstract

In the dynamic sphere of fashion, personalization plays a pivotal role in enhancing user experience. This study presents a novel fashion recommendation system that tailors clothing suggestions based on individual skin tones and body proportions, aiming to enhance personalization in fashion technology. The system employs a two-fold feature extraction approach: for skin tone analysis, it uses MediaPipe to detect and isolate the face in images, followed by conversion to the HSV color space to compute the average facial hue, discarding irrelevant facial features. For body proportion metrics, MoveNet Thunder extracts six pivotal keypoints to calculate ratios that define body shape. These extracted features from a substantial image dataset are input into k-nearest neighbors (KNN) and autoencoder algorithms to encode the complex relationships between body metrics and fashion items. Upon inputting an image into the system, it retrieves the top five most similar images from the dataset, leveraging distance metrics such as Euclidean, Manhattan, and Minkowski for comparison. Preliminary results indicate that the system can effectively suggest fashion items that complement the user's unique physical attributes, thereby enhancing the user experience in digital fashion platforms.

# Introduction

Fashion, a pervasive element of modern life, is no longer merely about clothing oneself; it has grown into a vibrant industry serving as a canvas for individual expression. As an integral part of our identities, people communicate their unique personalities and characteristics through their style and fashion choices. In an era where the pace of life is constantly accelerating, a staggering amount of time is devoted to selecting outfits - men invest roughly 13 minutes per day, which over a lifetime, amounts to four months. In comparison, women devote approximately 17 minutes daily, cumulating to an average of six months throughout their lives (Euse, 2016).

Just like the fast-paced world we live in; the fashion industry is an ever-evolving landscape with trends that shift at a breakneck speed. Moreover, many people encounter challenges in identifying the styles, colors, and fits that flatter their unique body proportions and skin tones. The rapid oscillation of these trends calls for an efficient optimization method to not only keep up with these changes but also to foresee upcoming fashion trends. In the pursuit of efficiency and precision, machine learning and deep learning algorithms emerge as crucial tools, automating and refining processes to save time. Data scientists have embarked on the task of compiling extensive datasets that can serve as the backbone for deep learning, thereby addressing these challenges.

Building on these ideas, this project aims to pioneer a leap forward in the fashion recommendation landscape. The objective is to construct a fashion recommending system that integrates these complex considerations of individual attributes - body proportions and skin tones - into its architecture, thereby providing tailored suggestions that truly compliment a user's unique physical attributes. In essence, the goal is to offer a more personal and fitting choice of clothes, making the daily task of selecting an outfit not just simpler, but also more enjoyable.

## **Research Background**

The concept of personal style is deeply rooted in individuality. Each person's style is unique to them, reflecting their personality, mood, and personal taste. This individuality makes it challenging for the fashion industry to provide personalized recommendations that cater to the unique needs of every customer. Traditional recommendation systems often use transaction history and popular trends to provide suggestions, but these methods don't consider the individual's unique physical characteristics (Bhure, et al., 2021).

In recent years, the power of artificial intelligence (AI) and machine learning has been harnessed to address this challenge. AI-powered recommendation systems have been created, providing personalized recommendations based on past purchases and browsing history (Chakraborty, et al., 2021). These systems have greatly improved the shopping experience, making it easier for individuals to find products they like.

However, these AI-based systems often overlook an important aspect: the physical characteristics of the individual. Skin tone and body proportions play a critical role in determining what styles and colors suit a person best. Research in color theory and aesthetic principles have proven this (Obeng, Danso, Omari, & Kuwornu-Adjaottor, 2018), but these factors are often not considered in current recommendation systems.

The idea of creating a fashion recommendation system that considers the individual's skin tone and body proportions is not new, but it has not been extensively explored (Wazhakar, et al., 2022). Most research has focused on style preferences and popular trends, while the importance of physical attributes has often been overlooked. This gap in research is the focus of this project.

By integrating skin tone and body proportions into the recommendation system, we aim to provide a more personalized shopping experience. By doing so, we hope to increase customer satisfaction and, in turn, boost sales and customer retention. This project will build upon existing research and incorporate underutilized data to create a more comprehensive and effective recommendation system.

## **Problem Statement**

Despite the advancements in technology and the growing reliance on artificial intelligence for personalized recommendations, a significant gap remains in the fashion industry. The current fashion recommendation systems largely depend on transaction history, style preferences, and popular trends (Deldjoo, et al., 2022). Unfortunately, these factors often result in a generalized approach that doesn't account for the individual's unique physical attributes such as skin tone and body proportions.

Skin tone and body proportions are critical elements when it comes to selecting flattering attire. Research has shown that suitable colors can enhance a person's complexion (Perrett, D. I. & Sprengelmeyer, R., 2021), while the appropriate fit can accentuate their body shape. However, existing recommendation systems largely neglect these crucial aspects, resulting in a lack of personalization.

The absence of a more personalized approach can lead to unsatisfactory shopping experiences and missed opportunities for businesses. Consumers may end up purchasing clothes that don't complement their physical attributes, resulting in lower customer satisfaction and possible product returns. This issue is exacerbated by the prevalent size-fit problem that leads to significant product returns and economic damage to companies (Deldjoo, et al., 2022).

Therefore, the problem lies in the fact that current fashion recommendation systems do not sufficiently address the need for personalized recommendations that consider individual skin tones and body proportions. There is a need for a more holistic approach that acknowledges these factors, leading to truly personalized and effective recommendations that enhance customer satisfaction and shopping experiences.

## Research Questions

In order to guide the development and assessment of the proposed fashion recommendation system, several research questions have been formulated. These questions serve as a compass, directing the focus of our study towards understanding how machine learning can be leveraged to interpret physical attributes from images and how this information can be used to design a personalized recommendation algorithm. They also aim to evaluate the effectiveness of our proposed system compared to existing solutions. The following key research questions will be explored in this study.

**Research Questions**

1. How can machine learning and deep learning algorithms be developed to accurately interpret skin tone and body proportions from images?
2. How can these interpreted attributes (skin-tone and body proportions) be incorporated into the design of a personalized recommendation algorithm?
3. How effective is the optimized algorithm in providing personalized fashion recommendations compared to existing systems?

## Aim and Objectives

**Aim:**

The primary aim of this project is to conceptualize, design, and optimize deep learning and machine learning algorithms to generate personalized fashion recommendations based on users' skin-tone and body proportions. This aim seeks to further explore the potential of artificial intelligence and machine learning in enhancing personalization within the fashion industry.

**Objectives:**

* Segment and extract the face portion from the images to determine skin tone by averaging the color values.
* Utilize pose estimation techniques to accurately extract body proportions from user-supplied images.
* Provide fashion recommendations that complement these physical attributes using Content-Based Filtering.

## Scope

The scope of this research project includes the following:

1. **Target Audience**: The research is primarily aimed at adult users, given their high engagement with fashion and established preferences. The system aims to serve users with diverse skin tones and body proportions. The target user attributes, namely skin tone and body proportions, will be extracted from a broad dataset of images that depict adults in various outfits.
2. **Feature Extraction**: The project aims to develop a system that can accurately extract and analyze individual skin tones and body proportions from user-provided images. Skin tones are obtained using MediaPipe to detect faces, remove irrelevant facial features, and convert the face to the HSV color spectrum to average the color. For body proportions, the MoveNet Thunder model is employed to extract six key points, which are then used to calculate body ratios.
3. **Dataset**: The system will be trained on a pre-compiled and diverse dataset of images that includes a wide range of skin tones and body types. The preparation and processing of this dataset are integral to the project, ensuring that the model is well-trained to make accurate recommendations.
4. **Algorithm Development**: The project involves the application of KNN and autoencoders, to analyze the extracted features.
5. **Recommendation Engine**: The core functionality of the system is to provide personalized fashion recommendations. By inputting an individual's image, the system will suggest the top five most similar garments from the dataset, with similarity calculated using distance metrics such as Euclidean, Manhattan, and Minkowski.

However, the project will not consider fashion preferences and trends or individual style statements. This limitation is due to the highly subjective and rapidly changing nature of fashion trends, which could complicate the algorithm development process. Additionally, the project will not focus on children's fashion due to the ethical complexities and sensitivity surrounding children's data.

This research project has been scoped to focus on skin tone and body proportions because these are consistent characteristics that can be effectively used to personalize fashion recommendations. Though the project is currently limited to adult users, the potential scalability of the algorithms and interface design could allow for applicability to a broader user base in the future. The research lays the groundwork for a more personalized, inclusive, and user-friendly fashion industry.

## Significance of the Research

1. **Personalization in Fashion:** This project aims to enhance the personalization capabilities of fashion recommendation systems. Personalization is increasingly critical in the fashion industry, with consumers seeking products that align with their unique preferences. By considering skin tone and body proportions, the system could make more accurate and relevant recommendations, potentially improving user satisfaction and engagement.
2. **Diversity and Inclusivity:** This project has the potential to contribute towards more diversity and inclusivity in the fashion industry. By including a wide range of skin tones and body proportions in its user model, the system can cater to users who might not be adequately served by existing fashion recommendation systems.
3. **Consumer Experience & Industry Impact:** A system that provides highly personalized clothing recommendations could potentially revolutionize the online shopping experience. It may not only increase customer satisfaction but also enhance customer loyalty, leading to repeat purchases and higher revenue for businesses.
4. **Basis for Future Research:** The methodologies and findings of this project could provide a basis for future research in this field. It might inspire further investigation into other factors that could be considered for personalization or the application of even more advanced AI techniques for fashion recommendation.

## Structure of Thesis

This thesis is systematically organized into distinct chapters, each serving a specific purpose within the broader context of the research. **Chapter 1:** **Introduction** sets the stage, providing an overview of the research background, defining the problem statement, and presenting the research questions, aims, objectives, and scope. It concludes with the significance of the research, highlighting its contribution to the field of fashion recommendation systems.

**Chapter 2:** **Literature Review** critically examines existing studies, outlining current knowledge and identifying gaps that this research seeks to fill. A detailed discussion and comparison of related works are presented, including a literature matrix in Appendix A for reference.

**Chapter 3: Methodology** details the procedural framework and the technical approaches employed in the study, covering data preparation, feature extraction techniques, and the algorithms used for the recommendation system. It also explains the rationale behind methodological choices and describes the workflow of the project.

**Chapter 4: Results and Discussion** presents the findings of the research, including an evaluation of the recommendation system's performance and its comparison with existing works. This chapter also discusses the implications of the results and acknowledges any limitations encountered during the study.

The thesis culminates in **Chapter 5:** **Conclusion and Future Works** chapter, which encapsulates the research's key takeaways and suggests directions for subsequent studies to build upon the findings of this research. The **References** section provides a comprehensive list of the scholarly works cited throughout the thesis.

Lastly, the appendices contain supplementary materials that support the thesis's research but are too voluminous or detailed to include in the main body of the text.

# Literature Review

This chapter embarks on a systematic exploration of the core methodologies that underpin the fashion recommendation system central to this study. It aims to furnish a foundational understanding of the algorithms employed for fashion recommending system. Furthermore, this review critically examines recent scholarly works that is related to this project.

## Introduction

In the realm of digital fashion retail, recommendation systems are not just a luxury but a necessity for customer satisfaction and business success. These systems have evolved from basic collaborative filtering to complex algorithms capable of tailoring suggestions to the individual's unique physical attributes and preferences. This literature review embarks on an exploration of the current methodologies and innovations in fashion recommendation systems, with a keen focus on user attributes.

Section 2.2 will distil the feature extraction process, an essential preliminary step for any recommendation system. It will delve into the components of skin tone extraction (Section 2.2.1) and pose estimation (Section 2.2.2), which together form the basis for understanding and quantifying the user's physical characteristics.

Following this, Section 2.3 will explore the frameworks of recommending systems, laying out the foundational concepts that guide the personalization of fashion suggestions.

The discussion will then narrow down to recommending algorithms in Section 2.4, with a special emphasis on unsupervised recommending algorithms (Section 2.4.1) that include KNN and autoencoders—algorithms that play a pivotal role in determining the closest matches to the user's extracted features without the need for labeled data.

Finally, the review will focus on fashion-based recommending systems (Section 2.5), culminating in an in-depth analysis of user attribute-based recommending systems (Section 2.5.1). This section will underscore the importance of integrating feature extraction with machine learning techniques to provide a seamless and personalized shopping experience.

Through this structured examination, we aim to highlight the critical role of advanced computational methods in the fashion domain, revealing how they can transform the online shopping landscape by providing precise and personalized clothing recommendations.

## Feature Extraction

This section delves into a comprehensive review of recent research on feature extraction methodologies with a particular focus on skin tone extraction and pose estimation, which are pivotal in calculating body proportions. The project includes an analysis of cutting-edge techniques utilized in skin tone analysis, examining the efficacy of various image processing tools such as MediaPipe in isolating and refining facial features to capture the skin tone. Additionally, we explore advancements in pose estimation algorithms, with an emphasis on the utilization of MoveNet Thunder, to accurately determine key bodily measurements. These methodologies are critically assessed to understand their roles in enhancing the accuracy and reliability of body proportion calculations within the scope of fashion recommendation systems.

### Skin Tone Extraction

Skin tone extraction is a methodical process employed in computer vision and image processing that involves the identification and isolation of the skin regions from digital images. The objective is to accurately determine the color of the skin, which can be a challenging task due to varying lighting conditions, shadows, and individual differences in pigmentation.

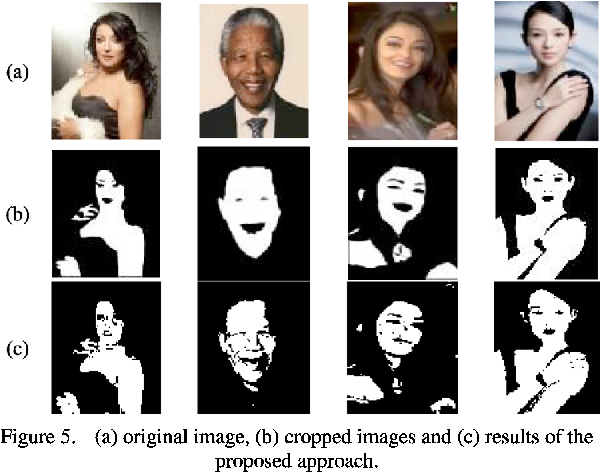


Figure 1. An illustration of the isolation of skin regions (Nada B. Ibrahim, M. Selim, & H. Zayed, 2012).

#### Challenges of Skin Tone Extraction

Skin tone extraction from digital images is a significant task, fraught with challenges that stem from both technical limitations and the intrinsic diversity of human skin tones. The most significant challenge in skin tone extraction is the variability of lighting conditions (M. A. Chyad, H. A. Alsattar, B. B. Zaidan, A. A. Zaidan, & Ghailan A. Al Shafeey, 2019). Skin tone can appear dramatically different under various light sources, such as fluorescent, incandescent, or natural sunlight, due to the different light spectra. This variability can lead to inconsistencies in the extracted skin tone, which may not truly reflect the individual’s natural coloring.

Moreover, human skin tones exhibit a wide range of colours and undertones, which can be challenging to capture accurately. The extraction algorithms must be robust enough to account for such diversity without bias towards any particular range of skin tones.

In addition, the presence of occlusions such as facial hair, cosmetics, or accessories can obscure the true skin area, while shadows cast by the facial features or external objects can create dark areas that do not represent the actual skin color, thereby complicating the extraction process (P. Kakumanu, S. Makrogiannis, & N. Bourbakis,, 2007).

Finally, the quality and resolution of the image also affect the precision of skin tone extraction. Lower resolution images may not provide sufficient detail for accurate color detection, while poor image quality can introduce noise that confounds the extraction algorithm.

#### Recent Studies on Skin Tone Extraction

The study under review (Diana, Adrian , & Radu , 2018) introduces an automated system for skin tone classification, tailored to support visagism applications such as virtual eyeglass try-ons. The novel classification framework is built upon two distinct methodologies: one that utilizes a Support Vector Machine (SVM) and another that employs Convolutional Neural Networks (CNN). The effectiveness of the system is rigorously benchmarked against current state-of-the-art techniques, revealing a marked improvement in accuracy, thus underscoring its potential applicability in the industry.

Central to both the SVM and CNN methodologies is the accurate extraction of the facial region from the input image. The research capitalizes on the well-established Viola-Jones algorithm for face detection. A region of interest (ROI), specifically beneath the eyes and above the center of the face, is then isolated to circumvent the interference of non-skin elements like hair or eyes, which could adversely affect classification outcomes. This strategic focus on relevant skin pixels forms the cornerstone of the system's robustness and precise skin tone classification.

In the SVM-based method, a moving window technique is applied to calculate color histograms of localized image patches within various color spaces. Post concatenation, these histograms undergo dimensionality reduction via Principal Component Analysis (PCA) before classification through an SVM. This method demonstrates an impressive accuracy of 86.67%, thereby emphasizing the critical role of color space selection and showcasing the enduring relevance of traditional machine learning algorithms in skin tone classification tasks.

Conversely, the CNN method, leveraging a finely tuned VGG-19 network, delivers a notable increase in classification accuracy, achieving 91.29%. This method surpasses the SVM approach and exhibits a marked reduction in misclassification between disparate skin tone extremes. This evidence the advanced discriminatory capabilities inherent in deep learning methodologies for feature extraction and classification within the domain of skin tone identification.

The comparative analysis delineated in the paper places the proposed methods in comparison with other skin tone classification solutions. It is observed that the CNN-based approach, in particular, provides a significant leap in accuracy over its predecessors, especially those with a comparable three-class skin tone categorization. Notably, the proposed methods demonstrate robustness without imposing constraints on the image acquisition conditions, rendering them more versatile for practical application scenarios.

In conclusion, the reviewed study presents a comprehensive skin tone classification system that operates without the necessity of color patterns or prior camera calibration. The combination of conventional machine learning (SVM) with advanced deep learning (CNN) techniques delineates the heightened efficacy of CNNs in the domain of skin tone classification. The research opens avenues for future enhancements and broader applicability in the rapidly evolving field of virtual try-on technology and personalized visagism solutions.

On the other hand, the research by (T. J. McBride, N. Vandayar, & K. J. Nixon, 2019) delves into an examination of various skin detection techniques, evaluating them on the metrics of accuracy and latency, utilizing the Pratheepan dataset for this purpose. Three primary techniques are extensively discussed: YCbCr Thresholding, RGB-H-CrCb Thresholding, and KNN Classification.

In the section on YCbCr Thresholding, the paper elaborates on how predefined threshold values within the YCbCr color space are utilized to differentiate between skin and non-skin pixels. It is noted that the algorithm performs effectively in detecting skin but encounters challenges when it comes to non-skin detection. Particularly, it struggles in very dark or very bright regions due to the omission of the luminance component. The performance of this technique is heavily influenced by the color of the background and the lighting conditions of the image, resulting in a level of inconsistency in accuracy.

The RGB-H-CrCb Thresholding section explores a more complex approach that combines the RGB, HSV, and YCbCr color spaces to enhance the classification of skin pixels. The algorithm showcased effectiveness in extracting skin regions from a majority of images. However, it was identified that the algorithm incorrectly classifies white or bright areas as skin and dark areas as non-skin. This misclassification leads to inaccuracies especially in images with white backgrounds or with individuals having darker skin tones. The paper describes how logical operations are employed to combine thresholding regions from the different color spaces, which minimizes the amount of skin detected in the final output, thereby maximizing non-skin classifications.

In discussing KNN Classification, the paper introduces the K-Nearest Neighbour Classification, a form of supervised machine learning algorithm, as a method to classify pixels as either skin or non-skin based on a training dataset. It is noted that the model demonstrates better performance in classifying non-skin pixels, a bias attributed to the training data which contains more non-skin pixel values. The algorithm faces challenges with skin-colored backgrounds and is also affected by the luminance of the image, resulting in inconsistent performance across different images.

In conclusion, the paper posits that while the RGB-H-CrCb thresholding algorithm recorded the highest average accuracy with the least standard deviation, it was still found to be inconsistent and unreliable under varying illumination and background conditions. All three algorithms demonstrated inconsistencies and high standard deviations in accuracy, signaling the limitations inherent in using predefined threshold values for skin detection in a general case model. The paper suggests exploring alternative solutions like adaptive thresholding, different machine learning algorithms, probabilistic models, or the creation of a more representative ground truth dataset to potentially improve the accuracy and consistency of skin detection for HGR applications.

### Pose Estimation

Pose estimation refers to the computational task in computer vision that involves detecting human figures in images or videos and inferring the spatial configuration of their body parts — typically locating key points on limbs, torso, and the head. The objective of pose estimation is to map the human anatomy within a two- or three-dimensional context, which allows for the analysis of body posture and movement.

#### Types of Pose Estimation

There are two primary types of pose estimation: **2D pose estimation**, which identifies the position of key points in the plane of the image (Haoming , Runyang , Sifan, Hao, & Fengcheng, 2022)., and **3D pose estimation**, which extends this analysis into the third dimension, providing a more complete representation of the body's pose in space (Jinbao , et al., 2021).

A screenshot of a computer

Description automatically generated

Figure 2. An illustration of the difference between 2D pose estimation and 3d pose estimation (Zatolokina, 2023)

Recent advancements in this domain have been propelled by deep learning, particularly through the use of Convolutional Neural Networks (CNNs), which can learn complex patterns in spatial data, making them well-suited for the task of detecting body joints and segments from visual inputs. The advent of large-scale annotated datasets has also significantly contributed to the progression of pose estimation algorithms.

#### Challenges of Pose Estimation

Pose estimation is an intricate task in computer vision, with occlusion presenting a prominent challenge. When key body parts are obscured by other limbs or objects within the image, accurately detecting and localizing them becomes complex. This occlusion often necessitates advance algorithms capable of inferring the position of obscured joints based on the visible anatomy and learned models of human posture (Zhao, Jianke, Jiajun, & Chun, 2015).

The translation of a three-dimensional world into two-dimensional images introduces ambiguity in visual perception. The same set of pixel configurations could represent different poses in 3D space, which adds a layer of complexity to the pose estimation task. This ambiguity means that the algorithms must often rely on contextual clues and probabilistic models to accurately interpret and reconstruct the pose (Shradha Dubey & Manish Dixit, 2021).

The complexity of backgrounds in images also adds to the challenge of pose estimation. In real-world scenarios, backgrounds can be highly varied and cluttered, making it difficult to distinguish the human figure from the background noise (W. Luo & J. Xue, 2023). Effective pose estimation algorithms must, therefore, be robust against a range of background conditions and able to isolate body parts even in complex scenes.

In summary, while pose estimation technology has made considerable strides in recent years, these challenges underscore the need for ongoing research and development to enhance the robustness, accuracy, and efficiency of these systems.

#### Recent Studies on Pose Estimation and its Applications

State-of-the-art pose estimation methods, such as those using deep learning architectures like MoveNet Thunder (Ronny & Na, 2021), address these challenges by employing robust feature extraction and inference mechanisms that are capable of handling the complexity and variability of human poses under a wide range of conditions. These models are trained on extensive datasets that encompass a diverse array of poses and scenarios to enhance their accuracy and generalizability. In applications like augmented reality, sports analytics, and human-computer interaction, pose estimation serves as a foundational technology that bridges the gap between human movement and machine understanding.

The paper by (Cristina , et al., 2022) presents an insightful analysis into the domain of pose estimation within computer vision, an area of growing interest due to its diverse applications from interactive gaming to aiding in physical therapy. The study focuses on comparing the efficacy of three prominent models: MoveNet, BlazePose, and PoseNet, specifically their performance metrics and adaptability to unregulated environments.

MoveNet emerges as the inital model for review. Structured upon the MobileNetV2 architecture, MoveNet is proficient in deducing 2D joint locations from RGB images. It presents two distinct variants: 'Lightning', which prioritizes swift real-time inference, and 'Thunder', which, albeit slower, delivers enhanced accuracy.

BlazePose, associated with the Mediapipe framework, operates via a dual-stage mechanism. Initial detection of personal Regions of Interest is succeeded by the computation of 2D and 3D keypoints. BlazePose is available in three models—'Lite', 'Heavy', and 'Full', each designed to optimize either speed, accuracy, or a harmonious blend of both.

PoseNet, the third model, built on the MobileNetV1 framework, is capable of simultaneously identifying the 2D joint locations of up to five individuals, although for the purpose of this study, the analysis is confined to single-person detection to ensure an equitable assessment.

The methodology implemented involves a two-stage evaluation process. The initial phase assesses the models against the COCO dataset to determine accuracy. Subsequently, the models' proficiency in handling 'in-the-wild' scenarios is appraised using a dataset of videos.

The findings reveal PoseNet as the frontrunner with 48.6% accuracy in predictions, effectively minimizing localization errors. MoveNet follows with a commendable 43.4%, while BlazePose falls behind with only 5.3% accuracy. Notably, in uncontrolled settings, MoveNet Lightning and BlazePose Lite demonstrated the highest processing speeds, positing them as ideal candidates for real-time applications.

In conclusion, despite PoseNet's superior accuracy, MoveNet is favored for its optimal blend of speed and precision, making it a dependable option for real-time pose estimation. The paper indicates that the evolution of video-based 3D body pose estimation is on the horizon, promising models with heightened accuracy and computational efficiency. For future research, the paper advocates for a nuanced understanding of user perception of accuracy, the creation of a tailor-made 3D dataset for 'in-the-wild' testing, and the integration of high-accuracy 3D pose estimation models in physical rehabilitation, potentially revolutionizing the field.

In the realm of online tailoring and garment fitting, the study (Daud , Manishankar , Sandeep, & Bikal , 2022) aims to enhance the precision of body measurements derived from images. The researchers developed a model that distinguishes between linear measurements (arm length, upper body length, etc.) and circular measurements (chest, waist, etc.) obtained from 2D images, typically captured by smartphone cameras. The paper delineates a two-fold approach: using pose estimation for linear measurements and edge detection in conjunction with keypoint detection for circular measurements.

The study utilized the Mediapipe framework to locate key points on the human body, enabling the calculation of linear measurements by identifying and connecting specific joints. The pose estimation method consisted of localizing body joints and configuring them into a valid human pose. It was imperative to process the images initially, standardizing them to a fixed size and converting them to grayscale to optimize the accuracy of the keypoints detected.

For circular measurements, the paper acknowledges the complexity of requiring both key points and body edges to ascertain the measurements of body parts like the chest or waist. The researchers employed the canny edge detection method to discern the body's edges from the image. They introduced a novel approach for calculating the ratio (R) of the distance in centimeters per pixel, differing from the standard pixel ratio used in other studies, to accommodate variations in image dimensions and distances from which images are taken.

The training datasets for both linear and circular measurements comprised images of individuals paired with their respective body measurements, taken manually. The inclusion of a variable kk, representing the average error or loss in measurement, was a critical factor in their model, which allowed for adjustments to be made to the generated measurements to enhance precision.

The analysis revealed that while the model was not entirely accurate, it held a promising accuracy rate, with 81.47% in linear and 72.18% in circular measurements, when considering a minimal margin of error. The researchers calculated the Root Mean Square Error (RMSE) to quantify the deviations between the predicted and actual measurements, suggesting that accuracy could be improved with a larger dataset to refine the kk factor.

In conclusion, this study presents a viable model for online tailoring applications, providing a method to obtain body measurements from images. The authors point out the potential for increased accuracy through the expansion of the dataset, which would fine-tune the error factor in the measurements. Despite resource limitations, the proposed model stands as a significant stride towards automating measurement extraction for the tailoring industry, particularly in the digital domain.

## Recommending System

Recommender systems are a subclass of information filtering systems that aim to predict the preferences or ratings that a user would give to a particular item. These systems are widely used in various domains to suggest products, movies, music, articles, or anything that a user might be interested in. The primary objective is to provide personalized recommendations to help users navigate through a large amount of available information (YuanZhe, 2022).

According to the study by (Deepjyoti & Mala, 2022), modern recommendations system can be broadly categorized into three different types, mainly being content-based recommender systems, collaborative recommender systems, and hybrid recommender systems. Content-based recommender systems work by understanding the content of items and a user's profile to make recommendations. They analyze item descriptions to identify key features and create a profile for each user based on their interactions with items. For example, if a user has watched a lot of action movies, a content-based system might recommend more movies of this genre. The main advantage of this type of system is that it can recommend novel items, even if they haven't been interacted with by other users. However, it may end up only recommending similar types of items, which could lead to a lack of diversity in recommendations. Moreover, collaborative recommender systems leverage the behavior and preferences of many users to make recommendations. They do not require knowledge about the content of the items. There are two main types: user-based, which finds users that are similar to the target user and recommends items those users liked; and item-based, which identifies items that are similar to the items the target user liked, based on other users’ interactions. One major advantage is that they can provide diverse recommendations, but a challenge they face is the cold-start problem where it's difficult to make recommendations for new users or new items with no interaction history. On the other hand, hybrid recommender systems combine elements of both content-based and collaborative systems to overcome the limitations of each. They can use various techniques to integrate the strengths of both types, for example by combining collaborative filtering and content-based filtering, or by using one to enhance the other. Hybrid systems have the potential to provide more accurate and diverse recommendations, solving problems like the cold-start issue to some extent and offering a richer set of recommendations by leveraging both item content and user interaction data.

A diagram of a person's connection

Description automatically generated with medium confidence

Figure 3. (a) Content-based Recommendation System (b) Collaborative Filtering-based Recommendation System. (YuanZhe, 2022).

## Recommending Algorithms

In a seminal study by Fei et al. (2019), a novel sequential recommendation model dubbed BERT4Rec was introduced, marking a significant stride in tackling sequential recommendation tasks. It is designed to handle sequential recommendation tasks by utilizing a deep bidirectional self-attention mechanism to model user behavior sequences. This approach aims to address some of the limitations found in traditional unidirectional models, which often fail to capture the full context of a user's interaction history. Unlike traditional methods that employ unidirectional models to encode user interactions from left to right into hidden representations, BERT4Rec identifies that such models can be sub-optimal as they restrict the power of hidden representation in users' behavior sequences and often assume a rigidly ordered sequence, which isn't always practical​. To overcome these identified limitations, BERT4Rec employs a deep bidirectional self-attention mechanism. This mechanism allows for the modeling of user behavior sequences in a manner that enables each item in a user's historical interactions to fuse information from both the left and right contexts. This bidirectional approach aims to provide a richer representation of user behaviors compared to unidirectional models​. In order to efficiently train the bidirectional model without information leakage, BERT4Rec adopts the Cloze task for sequential recommendation. This involves predicting random masked items in a sequence by jointly conditioning on their left and right context, thus learning a bidirectional representation model for making recommendations​. Moreover, BERT4Rec demonstrated superior performance over other sequential recommendation models like SASRec, establishing itself as a state-of-the-art baseline for sequential recommendations.

However, subsequent investigations have showed doubt on this superiority, underscoring the evolving and competitive nature of this domain, particularly as highlighted by (Aleksandr & Craig , 2022). In their discerning examination, the researchers embarked on a systematic review of studies comparing BERT4Rec and SASRec, only to discover a lack of consistency in the reported outcomes across these papers. To understand why this happened, an in-depth analysis of the available BERT4Rec implementations was undertaken. It was unveiled that a significant number of these implementations failed in replicating the initially reported results when trained with their default parameters. Moreover, it was revealed that the original implementation necessitated a considerably longer training duration compared to the default configuration to faithfully replicate the originally reported outcomes. This finding underscores that certain papers might have employed underfitted versions of BERT4Rec as baselines. In a constructive endeavour, the researchers proposed an alternative implementation of BERT4Rec, grounded in the Hugging Face Transformers library. This implementation, in a majority of instances, successfully replicated the originally reported results utilizing the default configuration parameters. It was further demonstrated that this implementation yielded comparable results to the most recent sequential recommendation models like NOVA-BERT and DuoRec. The researchers believe that this paper, along with the publicly accessible code, will serve as a valuable resource for fellow researchers, ensuring the utilization of appropriately trained baselines, thereby propelling the field in a forward trajectory.

On the other hand, the paper by (Xiangnan, et al.) introduced a general framework named Neural network-based Collaborative Filtering (NCF), aimed at harnessing the power of neural networks to model the user-item interaction patterns. Unlike conventional methods, NCF replaces the inner product operation with a neural architecture capable of learning an arbitrary function from data, thereby offering a more expressive and flexible model. Within the NCF framework, the authors introduced three distinct models: Generalized Matrix Factorization (GMF), Multi-Layer Perceptron (MLP), and a fusion of the two termed Neural Matrix Factorization (NeuMF). GMF extends the traditional matrix factorization by replacing the inner product operation with a neural architecture, thus allowing for a more flexible representation and potential non-linear interactions between user and item latent factors. Whereas MLP model, on the other hand, learns interactions between user and item latent features by applying a series of non-linear transformations. This model enhances the expressiveness and flexibility in capturing complex user-item interactions beyond the linear interactions captured by GMF. Moreover, NeuMF combines the strengths of both GMF and MLP by concatenating their last hidden layers to learn complex user-item interactions. This fusion model embodies both the linearity of GMF and the non-linearity of MLP, providing a richer representation of user-item interactions. The empirical evaluation carried out by (Xiangnan, et al.) on two real-world datasets showcased the superiority of NCF over state-of-the-art methods. Notably, the experiments evidenced that employing deeper layers of neural networks leads to better recommendation performance, underscoring the potential of deep learning in collaborative filtering tasks.

Furthermore, the NCF framework was presented as a generic model, capable of expressing and generalizing matrix factorization under its structure, thus bridging the gap between traditional collaborative filtering methods and neural network-based models.

This research is pivotal as it not only challenges the traditional paradigms of collaborative filtering but also lays a foundational framework for future explorations into the integration of deep learning techniques in recommender systems. The introduction of NCF marks a significant step towards the evolution of more expressive, accurate, and robust collaborative filtering models, thereby contributing to the broader discourse on enhancing personalization and recommendation quality in modern digital platforms.

### Unsupervised Recommending Algorithms

Unsupervised Recommender Algorithms refer to a class of algorithms used in recommendation systems that operate without the need for labelled training data. Unlike supervised learning, which relies on historical data with predefined outputs to make predictions, unsupervised learning algorithms identify patterns, correlations, and structures within the data itself to make recommendations. In particular to the project, the focus is solely on the KNN and autoencoder algorithms.

#### KNN

The K-Nearest Neighbors (KNN) algorithm, traditionally known in the realm of supervised learning for its simplicity and efficacy, finds a versatile application in unsupervised learning tasks, including clustering and recommendation systems (Jingwen , Weixing , & Niancai , 2018). While KNN is inherently a type of instance-based learning, primarily utilized for classification and regression by identifying the 'k' nearest data points to the query instance, it can be adapted for unsupervised learning, particularly in identifying natural groupings of data.

In unsupervised learning, the aim is to discern intrinsic structures within the data without pre-assigned labels. KNN contributes to this by measuring distances between data points (commonly using Euclidean distance but not limited to it) and identifying the 'k' nearest neighbors for each point. In the context of clustering, this creates a network or graph of nodes, where each node is connected to its 'k' closest neighbors. By interpreting this graph, one can define connected components or clusters, with the idea that a cluster comprises points that are mutually nearest neighbors, reflecting a higher degree of similarity or closeness.

For recommendation systems, this clustering approach becomes particularly valuable. It can be utilized to group similar items, which, in turn, can enhance the accuracy of item-based recommendations. If a user expresses a preference for a particular item, other items within the same cluster can be recommended, assuming they share attributes or features that define the cluster's coherence. Similarly, clustering users based on their behavior or attributes allows for user-based recommendations, where a user may receive suggestions informed by the preferences of others within the same cluster.

Moreover, KNN serves a pivotal role in recommendation systems, not for clustering but for conducting similarity searches. It calculates distances between feature vectors to identify and retrieve items most closely aligning with a user's input (Leonardo et al., 2018). This mechanism is comparable to searching for books in a well-ordered library based on a reader's past selections rather than by genre classification. In this way, KNN acts as a navigational tool in a multidimensional feature space, guiding users to fashion items that are nearest to their unique combination of features, thereby personalizing the recommendation experience.

A diagram of a network

Description automatically generated

Figure 4. This figure demonstrates the application of the k-Nearest Neighbors (kNN) algorithm for retrieving items based on similarity in a feature space. The red dot symbolizes the query input, for which the system is tasked with finding similar items. The surrounding blue dots depict the various data points within the dataset, each positioned according to its feature values. Those blue dots that are directly connected to the red dot signify the five closest neighbours to the query input—determined by setting K=5K=5—illustrating the data points that bear the smallest distances and, consequently, the greatest similarity to the query.

In essence, while KNN's role in supervised learning is well-established, its utility in unsupervised learning, particularly in the clustering and recommendation domains, is equally significant. It offers a method to uncover the latent structures within data, which is foundational for any system aiming to curate and recommend based on underlying patterns of similarity.

In the study by (Prof.Shivganga, Jayesh, Prajwal, Harshal, & Shusovan, 2020), the authors presented an innovative book recommendation system utilizing the K-Nearest Neighbors (KNN) algorithm in conjunction with cosine similarity for collaborative filtering. Addressing the challenge of handling a large dataset containing approximately 600,000 entries from the Institute for Information Freiburg, the methodology involved refining the data by excluding books rated by fewer than 100 people and users who provided fewer than 200 ratings. This filtration was implemented without bias towards the age of the raters, although it was noted that the majority of ratings were contributed by users between the ages of 20 and 40.

The recommendation system was designed to process user ratings, which ranged from 0 to 10, by creating a matrix that correlates user ratings with book titles. By treating each cell in the matrix as a vector, the authors applied cosine similarity to calculate the dot product between vectors, thus establishing similarity. The KNN algorithm was then employed to determine the Euclidean distance between these vectors based on the chosen 'k' value.

The results demonstrated that collaborative filtering, particularly when employing a substantial dataset, proved more efficient and required fewer computational resources than content-based filtering approaches. The system was deployed as a web application for a bookstore, using standard web technologies and a refined dataset to improve recommendation accuracy.

On the other hand, the study by (Alhamza , Rozmie , Mosleh , & Mohammed , 2016) provides a comparative analysis of three unsupervised machine learning algorithms: K-means clustering, K-Nearest Neighbor (KNN), and Expectation Maximization (EM), in the context of network traffic classification. Using the Weka software and the Moore dataset, the study evaluates the performance of these classifiers based on classification accuracy, speed, and memory consumption—key metrics for real-time and online classification environments.

The findings indicate that KNN, with a configuration of three neighbors, offers the highest classification accuracy at 98%, and is also the most memory-efficient among the three methods. However, it is suggested that KNN could potentially become less efficient as the number of neighbors increases. On the other hand, K-means stands out for its speed, taking only 60 seconds for total processing, making it suitable for real-time applications, though it lags in accuracy and memory efficiency. EM ranks lowest due to its high computational cost, resulting in significant memory and time usage.

The study concludes that while KNN excels in accuracy and memory usage, optimizing its processing time could make it more suitable for real-time applications. Conversely, improving the accuracy and reducing memory consumption of K-means could render it more effective. The authors suggest further research to optimize both algorithms for use with large datasets in real-time and online settings.

#### Autoencoder

An autoencoder is a type of artificial neural network used to learn efficient coding of unlabeled data, typically for the purpose of dimensionality reduction or feature learning. It is designed to reconstruct its own inputs, which forces the network to try to learn the most important features in the data in a lower-dimensional space (Dor , Noam , & Raja , 2020).

The architecture of an autoencoder is typically composed of two main parts: the encoder and the decoder. The encoder compresses the input into a latent-space representation, and the decoder reconstructs the input from this representation. The latent space is a lower-dimensional space of features that the network has deemed the most informative for reconstructing the input data (Shuangshuang & Wei, 2023). By training the network to minimize the difference between the input and its reconstruction, the autoencoder learns a representation of the data.

In the context of unsupervised learning, autoencoders are particularly useful. Since they do not require labels to learn from the data, they can make sense of the data by identifying patterns and reducing dimensionality. This capability is beneficial in clustering and recommendation systems (Pengzhi, Yan, & Jianqiang, 2023). For clustering, the latent space learned by the autoencoder can reveal the intrinsic structure of the data. Once the data is encoded, traditional clustering algorithms like K-means can be applied more effectively in latent space. This is because the autoencoder, by capturing the most salient features, can often remove noise and redundancy in the data, resulting in cleaner and more distinct cluster separations.

In recommendation systems, autoencoders can be used to predict user preferences. By learning to compress user-item interaction matrices into lower-dimensional representations, autoencoders can capture the underlying factors that determine how users might rate or prefer certain items (Diana , Sofia, António , & José , 2020). This compressed knowledge can then be decoded to predict missing entries in the user-item matrix, effectively recommending new items to users based on their learned preferences.

In both cases, autoencoders serve as powerful unsupervised learning models that can uncover hidden structures in the data, which are instrumental for clustering or making accurate recommendations without the need for labeled training data.

## Fashion Based Recommending System

Fashion Recommending System (FRS) can be characterized as a method of aligning features of fashion items with users or consumers based on particular matching guidelines (Samit , Md. Saiful, Naimur , & Manik , 2021). According to the research conducted by (Shaghayegh , 2021), a fashion recommendation system principally encompasses four tasks: recommending similar or identical items (item retrieval), suggesting complementary items, proposing outfit combinations, and recommending a capsule wardrobe.

A screenshot of a computer

Description automatically generated

Figure 5. An illustration that encapsulates the primary functions of fashion recommendation system. (a) item retrieval, similar items are recommended based on the features of the input image (Samantha Jackson, 2019). (b) suggesting complementary items, complementary items are generated with GAN based on the input image (top) (Sudhir & Mithun , 2019). (c) proposing outfit combinations (Huang, 2021). (d) recommending a capsule wardrobe (Ying & Tao, 2017).

The research by (YASHAR , et al., 2022) highlights several key challenges in the domain of fashion recommendation systems. First, accurately representing fashion items is difficult due to the sparsity of data, with emerging solutions utilizing detailed visual, textual, and video inputs that demand large datasets for effective generalization. Second, predicting whether fashion items are compatible, essential for creating outfits, involves analyzing co-purchase data, designer insights, and social media trends. Personalization adds another layer of complexity, requiring the integration of variables such as location, occasion, and personal style, along with considerations of fit and body shape. Additionally, the need for systems that can provide interpretable and transparent recommendations is emphasized to build user trust. Lastly, the paper touches on the challenge of detecting and predicting fashion trends, which is multifaceted due to influences like seasonal changes, geographic locations, and the dynamic nature of style, with social media serving as a valuable data source for this analysis.

Research conducted by (Jaechoon , Seolhwa , Chanhee , & Dongyub, 2020) showcases a deep learning-based fashion product retrieval model. This study delves into the development of intelligent fashion techniques, introducing two key models: the Sketch-Product Fashion Retrieval Model and the Vector-Based User Preferred Fashion Recommendation Model.

The Sketch-Product Fashion Retrieval Model seeks to overcome the limitations of text-based search by enabling users to use images or sketches for product search. Initially focusing on image-based retrieval, it later incorporated a sketch-based retrieval method. The model employs deep learning to upscale a user's sketch to an image level and performs vector-based image comparison to find similar products. Evaluation metrics such as “Precision at 5” were used to measure its performance, with the image-based retrieval sub-model scoring a "Precision at 5" of 0.774, and the sketch-based one scoring 0.445, indicating a significant improvement over a baseline random model.

On the other hand, the Vector-Based User Preferred Fashion Recommendation Model aims to provide personalized fashion recommendations by profiling user preferences. Initially pre-trained with professional profiling data, it later adapts to individual user preferences. Utilizing a Deep Neural Network (DNN), the model learns from a dataset of paired top and bottom apparel items labeled according to user preference. The model was tested on 10,000 cases of menswear data from a web shopping mall, showing positive performance in recommending preferred fashion choices.

In conclusion, the experiments conducted on men's apparel e-commerce data demonstrated the potential of these models in enhancing the efficiency of fashion product search and recommendations. The Sketch-Product Fashion Retrieval Model notably improved the search process, while the Vector-Based User Preferred Fashion Recommendation Model showed promise in delivering personalized fashion recommendations. The study suggests that implementing these models on online shopping platforms could enhance user satisfaction and sales revenue. Future work will focus on real-world implementation and analyzing consumer purchasing patterns on actual shopping mall sites.

According to the study by (Soham , Krish , Harsh , & Dr. Suvarna , 2022), the researchers presented a WebApp that simplifies finding matching outfits. Users upload an image, which is processed using a Convolutional Neural Network (CNN) to resize and segment it, then convert it to a vector form for feature extraction. The system then searches for similar items within a dataset, providing shopping links for the matches. The Agile methodology was chosen for its efficiency in the development process, allowing for rapid, iterative updates with a focus on continuous testing and quality. The system is composed of an Input module for image upload, an API module to handle requests and connect with the machine learning model, and an ML module that processes the image and finds similar outfits using CNN and similarity algorithms. The final product offers a user-friendly interface for uploading images and instantly receiving fashion recommendations with options to directly purchase the recommended items online. The model aims to alleviate the difficulty of finding the right design and color in fashion, providing quick and tailored outfit suggestions.

In the research of (Aneesh K, P V Rohith , Sai , & Archana, 2022), the researchers introduced a three-stage fashion recommendation system that utilizes Convolutional Neural Networks (CNNs) for fashion recommendations. Initially, a CNN model discerns the color of clothing from an image, categorizing it into one of twelve possible color labels. Subsequently, a separate CNN model identifies the clothing type, distinguishing among shirts, shoes, pants, and t-shirts. Following the identification process, the system stores these details and activates a recommendation algorithm. This algorithm, which can optionally incorporate an item from the user's wardrobe as input, suggests outfit combinations by analyzing type, color, and occasion. It uses fashion matrices for different occasions to compute the most complementary color combinations for pants, shirts, and shoes. The algorithm then generates a list of top outfit matches, visually represents them with images from the user’s wardrobe, and continues to iterate through combinations, providing a dynamic array of fashion choices tailored to the user's needs and existing wardrobe, thus facilitating a more personalized and efficient dressing experience.

In the analysis by (Guan, C., Qin, S., Ling, W., & Long, Y., 2018) , the researchers presented a knowledge-based apparel recommendation system that relies on a complex communication network model, integrating both tangible visual elements and intangible meaning layers to understand and recommend fashion items. The system recognizes apparel signs such as line, shape, and color, which constitute the visual aspects of clothing. These signs are connected to denotation meanings (literal attributes like lapel shapes) and connotation meanings (subjective perceptions like formality or attractiveness).

The data used by the system is classified into two datasets: ATTRIBUTE, which captures obvious clothing features, and MEANING, which encompasses the connotative aspects contributed by fashion professionals. These datasets are evaluated for reliability and validity through various statistical and content-based assessments.

For the implementation, a multi-task Convolutional Neural Network (CNN) is designed to extract image-based features, classify style genres, and predict body shapes. The CNN's predictions on attributes are then used to train a separate Meaning Prediction model using two classifiers: Support Vector Machine (SVM) and a novel Later Kernel Fusion (LKF). The LKF classifier shows improved prediction accuracy over the SVM.

The results of the models demonstrate an effective system for recognizing detailed clothing features and styles, matching or exceeding the capabilities of trained professionals, with an average prediction accuracy of 88.1% using the LKF classifier. While currently tailored for menswear, future expansion could include larger and more diverse datasets, additional apparel categories, and a broader range of style and occasion-based meanings.

As found in the study by (M Sridevi, N ManikyaArun, M Sheshikala, & Sudarshan E, 2020), the researchers presented a novel fashion recommendation system that leverages a Convolutional Neural Network (CNN) informed by transfer learning from the ResNet50 model. The system is trained using a fashion dataset, with additional layers incorporated into the ResNet50 architecture to tailor it for fashion item recognition. Fashion items from Rent the Runway's inventory are processed through this neural network to create a database of item embeddings. Recommendations are generated using the Annoy library to find the closest matches to a user-uploaded image, based on visual similarity through Cosine Similarity measures. The system's efficacy is demonstrated by high accuracy and Fbeta scores after being trained and validated on the DeepFashion dataset. The robustness of the model is further confirmed through successful tests with a diverse set of images sourced from the internet and real-world captures. This system simplifies and enhances the online shopping experience by allowing customers to receive fashion product recommendations that closely match the style of images they upload.

Building on the research by (Sudhir & Mithun , 2019), which utilized an enhanced conditional Generative Adversarial Network (c+GAN) to create matching bottom items for given top garments, the authors developed a comprehensive approach to fashion recommendation. Their methodology began with the collection of a large dataset (100k images) from Bing images, focused on the most recent fashion trends. This dataset was refined through several stages: using Faster R-CNN to identify full-body images, segmenting these into tops and bottoms with crowd-sourced help, and then filtering down to a workable set of 35,000 image pairs.

The novelty of their approach lies in the training of the c+GAN model, which was adapted to better handle the variability and complexity of fashion images. They incorporated Mean Squared Error (MSE) and perceptual loss into the GAN's generator to improve the visual quality and detail of generated items. Additionally, they employed clustering techniques to manage the variability in pose and style, facilitating more effective training batches for the GAN.

Objective evaluations using Search Engine Criteria (SEC) for quality and color diversity entropy for diversity showed that these modifications led to more accurate and varied recommendations. The end product of this process is not just theoretical; it has been applied in a real-world context, powering the 'Goes Well With' feature in Bing's Shopping vertical, demonstrating the practical application and potential commercial value of this research.

To build upon this foundation, future research could explore further enhancements to the c+GAN architecture, such as incorporating three-dimensional modeling to understand the draping of garments on various body types. Additionally, integrating user feedback loops could refine recommendations based on personal preferences and historical data, potentially increasing the personalization of fashion item recommendations. Finally, expanding the dataset to include a broader range of fashion categories and styles could ensure the model's versatility and its ability to adapt to the fast-paced changes in fashion trends.

Moreover, as found in the study by (Wang-Cheng , Chen Fang, Zhaowen , & Julian , 2017), the researchers introduced a multifaceted system designed to revolutionize fashion recommendations. This system merges a visually aware recommender, known as DVBPR, with an innovative image generation component, harnessing the capabilities of deep learning for a personalized user experience. The DVBPR, an end-to-end deep Bayesian personalized ranking method, harnesses visual features from fashion items using a Convolutional Neural Network (CNN) based on the CNN-F architecture. It predicts user preferences through a Matrix Factorization approach enhanced by visual features and personalized latent factors, extracted directly from images.

The system’s CNN architecture is meticulously designed with multiple layers to efficiently train and capture intricate visual details. It operates on the principle of Bayesian Personalized Ranking (BPR), aiming to curate a user-specific ranking of items by analyzing implicit feedback like clicks or purchases. The BPR framework optimizes the model by ranking observed items higher than non-observed ones for each user.

Complementing the recommender is the image generation segment, powered by Generative Adversarial Networks (GANs), specifically conditional GANs. These GANs craft images from a noise vector conditioned on product categories, thus enabling the generation of new fashion items that resonate with the user's style preferences. The discriminator component of the GAN evaluates the authenticity of images, distinguishing between those that are part of the training set and those synthetically produced by the generator.

The system extends its utility to personalized design, where it can create images that not only appeal to the user’s style but also push the boundaries of fashion by generating novel design elements. This process draws parallels to activation maximization in neural networks, aiming to amplify the user’s preference value in the generated designs.

(Wang-Cheng , Chen Fang, Zhaowen , & Julian , 2017) system was rigorously tested through quantitative methods, using metrics like AUC, and qualitative assessments for image synthesis and design capabilities. When pitted against several baseline methods, the DVBPR stood out, offering superior recommendation performance. This study presents a cutting-edge blend of recommendation and generative models, setting a new precedent for personalized fashion recommendation and automated design.

### User Attribute Based Fashion Recommending System

As demonstrated in the research by (Atharv , Kunal , Manav , & Neha , 2020), The system is designed to be user-centric, offering an interface through which users can interact with it. Users have the ability to upload images of clothing into a digital wardrobe stored in their mobile device's cache memory. In addition to these visual inputs, the system also requires users to input personal attributes such as age, gender, and skin tone, along with their preferences, including favorite colors and brands. These inputs are essential for the system to provide customized results.

The sophisticated interface comes loaded with features that, upon activation by the user, prompt the system to execute specific backend functions to yield desired outcomes. The automated process involves detecting clothing from uploaded images, cropping, aligning, and then performing segmentation and categorization through specialized modules that have been trained and validated on a dataset.

The system employs Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to extract features from the images, which are then converted into a feature vector format. These vectors are crucial for the functioning of various algorithms within the system. There are specially designed matching and recommendation models that sift through the data to provide the user with the best possible combinations and suggestions that align with their preferences. These models work in tandem with the preference model, which factors in user attributes to deliver personalized results.

Conversely, the research conducted by (Seema , et al.) introduced a fashion recommendation system tailored to body shape. This innovative system employs Bayou’s equation to determine the customer's body shape and categorizes them into one of eight distinct body types. The dataset utilized in their study was meticulously collected in-house, comprising exclusively women's apparel images. This approach enabled the deployment of pre-trained deep learning models, such as the Xception model, which emerged as the most accurate in aligning fashion items with the corresponding body types. By harnessing these advanced models, the system demonstrated remarkable proficiency in suggesting suitable garments, effectively catering to the nuances of different female profiles. For instance, the recommendation algorithm was adept at identifying and proposing long-sleeved shirts for inverted body types, flowy tunics for those with an apple shape, and form-fitting skinny jeans for hourglass figures. This precision in recommendation not only enhances the shopping experience for the consumer but also offers a potent tool for retailers in the fashion industry to optimize their offerings.

On the other hand, in the research of (Digant, Anushree , Monali , Shivani , & Tabassum , 2019), the researchers presented a novel approach to fashion recommendation that uniquely incorporates user-specific parameters such as skin tone, occasion, current trends, gender, and size to curate personalized clothing suggestions. Initially, the user is prompted to upload a photograph, which the system analyzes using a k-means clustering algorithm to accurately extract the skin tone. This skin tone extraction is crucial as it informs the subsequent decision tree algorithm, which then correlates the user's occasion and clothing size to generate suitable outfit selections from an extensive dataset.

For skin tone detection, the system takes the user's image, applies face detection, and then uses k-means clustering to identify the predominant skin tone, which matches against predefined skin tone shades to classify the user's skin tone. Following this, outfit classification is achieved by associating skin tone classes with colors that are either recommended or not recommended for that skin tone. The system categorizes outfits into classes where some classes represent clothes not recommended for certain skin tones, while one class includes universally suitable options.

The underlying decision tree algorithm takes into account the determined skin tone, along with the occasion and the user's size, to showcase clothing that not only fits well but also harmonizes with the user's natural coloring. It effectively filters out any garments that fall within the skin tone shades categorized as unflattering for the individual.

A meticulously assembled manual dataset of 1030 outfits, stratified by occasion, gender, and size, serves as the backbone for the simulation of this recommendation process. The system offers a selection of the top 10 outfits to the user, incorporating their feedback into the dataset, which is essential for the continuous learning and improvement of the algorithm.

The system's performance was rigorously evaluated on both the accuracy of the skin tone detection and the efficiency of the execution time. Although the results showed promise, a potential limitation of the study is the lack of clarity on the metrics used for evaluating the recommendation's effectiveness. Furthermore, the evaluation could benefit from a larger, more diverse dataset to ensure robustness and reliability across various demographics. As future work, the researchers aim to expand the dataset and refine the algorithm to minimize any discrepancies, thereby improving the precision of the recommendations.

In the research of (Reeta, Anisha , Tejashri , & Siddesh , 2021), the researchers presented a sophisticated outfit recommendation system named “Pocket Fashionista”. It composed of multiple modules, each designed to refine the fashion selection process with a touch of personalization. At its core, the system features a Skin Tone Detection and Classification module where users upload a photograph, and the system, using OpenCV, classifies their skin tone into one of five predefined categories. This classification is critical as it informs the outfit color recommendations. The system employs skin segmentation through RGB, HSV, and YCbCr color models, followed by a skin thresholding algorithm to differentiate skin from non-skin regions accurately.

Following skin tone identification, the system recommends outfit colors in the Outfit Color Recommendations module. Leveraging a dataset that includes men's formal shirts, the module fetches clothing images that match the user’s skin tone. These recommendations are visually presented to the user, enabling an interactive selection process. Another innovative module is the Weather-based Outfit Recommendations, which uses the DeepFashion dataset in conjunction with Convolutional Neural Networks (CNNs) and transfer learning techniques from models like ResNet. This module is adept at suggesting seasonally appropriate attire by processing image data to categorize outfits for different weather conditions.

A standout feature of the system is the Virtual Try-On module, which allows users to virtually wear selected outfits using a live video stream. This is made possible through real-time image processing and object detection via the OpenCV library, which superimposes the chosen clothing onto the user’s image. Complementing this is the Similar Outfit Recommendations module, which employs a content-based recommendation approach. By extracting features from the Feidegger dataset using the pretrained VGG16 model, it generates a similarity score matrix that the system uses to suggest outfits resembling the user's current selection.

Despite the innovative approach and features of "Pocket Fashionista," the system's efficacy is contingent on several factors that could be considered limitations. For instance, the accuracy of the skin tone classification and the subsequent color recommendations might vary with the quality and lighting of the user-uploaded photos. The current dataset, although extensive, may not encompass the full spectrum of global fashion trends and personal preferences, which could limit the diversity of the recommendations. Furthermore, the recommendation algorithms rely heavily on static datasets, and their performance may not reflect real-time changes in fashion trends. As the system expands, considerations for a more dynamic dataset and algorithms that can adapt to evolving fashion trends would be beneficial in maintaining relevance and accuracy.

## Discussion and Comparison of Related Works

The integration of AI in fashion recommendation systems has marked a significant transformation in how consumers interact with fashion e-commerce platforms. The current landscape of research, as illustrated by the array of studies summarized earlier, showcases a shift towards highly personalized and user-centric models. As Shaghayegh (2021) delineates, these systems are tasked with a multi-faceted objective: item retrieval, complementary item suggestions, complete outfit composition, and capsule wardrobe creation, all tailored to the user's preferences and physical attributes.

A closer examination of the methods reveals that while many studies employ Convolutional Neural Networks (CNNs), the application and fine-tuning of these models vary considerably. For instance, the approach by Aneesh K, et al. (2022) to categorize clothing into color labels contrasts with the knowledge-based system presented by Guan, C., et al. (2018), which integrates both visual elements and intangible meanings. This diversity in methodologies underscores the complexity of fashion recommendation and the need for multifaceted systems that can navigate between the tangible attributes of clothing and the abstract personal preferences of users.

The use of CNNs informed by transfer learning, as demonstrated by M Sridevi, et al. (2020), highlights the potential for leveraging pre-existing models to enhance recommendation accuracy. However, the limitation to specific inventories, such as Rent the Runway, poses questions regarding the scalability and adaptability of such systems to a wider market.

In contrast, the studies by Sudhir & Mithun (2019) and Wang-Cheng, et al. (2017) explore the generative capabilities of AI, using c+GANs and GANs, respectively, to not only recommend but also create new fashion items. This generative approach suggests a growing yet promising avenue for AI in not only responding to but also shaping consumer fashion choices.

Notably, personalization extends beyond mere style preferences, as indicated by Digant, et al. (2019), who factor in skin tone and body type into their recommendations. This consideration of physical attributes is a critical dimension that many systems overlook, yet it is crucial for ensuring that recommendations are not only aesthetically pleasing but also flattering to the individual user.

The datasets utilized across studies are as varied as their methodologies, ranging from proprietary collections to public datasets like DeepFashion. The choice of dataset heavily influences the system's performance, as seen in the work of Seema, et al., where an in-house dataset curated for women’s, apparel facilitated highly customized recommendations.

While the advancements are notable, the field is not without challenges. The studies reveal common limitations such as a focus on narrow clothing categories, the need for larger and more diverse datasets, and a lack of real-time adaptability to changing fashion trends. These limitations suggest that future work should aim for broader applicability and responsiveness to the fast-paced nature of the fashion industry.

In summation, the related works present a convergence towards AI-driven personalization in fashion recommendation systems, each contributing a unique piece to the complex puzzle. The journey from static product suggestions to dynamic, personalized wardrobe curation is ongoing, with each study paving the way for more intelligent, adaptive, and individual-centric fashion technologies.

A comprehensive synthesis of the reviewed literature is organized into a literature matrix, which can be found in Appendix A. It is important to acknowledge that the referenced studies serve as a contextual framework rather than a basis for direct comparison, due to the heterogeneity in datasets employed and methodologies applied across the research. These variations inherently yield distinct outcomes and insights. Therefore, the matrix is intended to provide a structured overview that found an understanding of the scope and diversity within the field, rather than to serve as a tool for direct empirical comparison.

# Methodology

This chapter meticulously delineates the methods and strategies employed to fulfil the aim and objectives outlined in the previous sections. It serves as a roadmap guiding the technical and analytical procedures undertaken throughout the course of this project. The methodologies are articulated in a structured manner to provide a clear understanding of the step-by-step approach adopted, the tools and techniques utilized, and the rationale behind choosing these methods.

## Workflow of the Project

The project initiated with feature extraction, focusing on two key attributes: skin tone and body proportions. This initial phase involves analyzing images and quantifying the mentioned features, which are then methodically compiled into a CSV file for subsequent processing.

Post extraction, the gathered features serve as inputs for the k-Nearest Neighbors (kNN) algorithm. This machine learning technique is employed to measure the similarity between different images based on the extracted features. The kNN operates by locating the 'k' closest neighbors in the feature space, providing a basis for image recommendations.

To enhance the robustness of the system, the data undergoes normalization to ensure that all features contribute equally to the similarity calculations. This step is crucial for maintaining the integrity of the kNN results, especially when dealing with features that vary in scale and range.

In parallel, an autoencoder neural network is trained to learn a compressed representation of the data. This network consists of two parts: an encoder that reduces the data to a lower-dimensional space and a decoder that reconstructs the data from the reduced form. The autoencoder is optimized using a loss function that minimizes the difference between the original and reconstructed data, with the aim of capturing the most salient features in the encoded representation.

After training, the encoded features from the autoencoder provide an alternative similarity measure. By computing distances in the encoded space, we can capture more nuanced patterns that may not be evident in the raw feature space used by the kNN.

The system provides recommendations by comparing the feature vector of an input image against the dataset. It utilizes both kNN and autoencoder-derived distances to offer a comprehensive set of similar images. The recommendations are then ranked, and the top results are presented to the user.

To facilitate user interaction, a graphical user interface (GUI) is developed using Tkinter. This interface allows users to upload an image and receive recommendations with ease. The GUI displays the recommended images along with their respective distances from the input image, offering insight into their similarity levels.

Finally, the system's effectiveness is gauged by assessing the relevance of the recommendations provided. This assessment can be based on objective metrics, such as distance measures, or subjective user feedback. Continuous refinements are made to improve the system, with the goal of delivering the most accurate and satisfactory recommendations to the user.

Further elaboration on each segment of the methodology is provided in the following sections, delineated from Section 9.2, "Data Preparation", to Section 9.4.2, "Autoencoder". A schematic representation summarizing the workflow is presented in the subsequent figure.

A screenshot of a computer screen

Description automatically generated

Figure . Overview of the Project's Workflow

## Data Preparation

The dataset utilized in this project was sourced from the work of Ziwei, Ping, Qiu, XiaoGang, and Xiaoou (2016). Permissions were duly obtained to access and utilize this dataset for the purposes of this study.

The dataset, known as DeepFashion, is a robust and diverse collection encompassing over 800,000 fashion images, illustrating a broad range from well-posed shop images to unconstrained consumer photos, distinguishing it as the most expansive dataset for visual fashion analysis thus far. It is elaborately annotated, with each image categorized into one of 50 distinct classes, further enriched with 1,000 descriptive attributes, and marked with bounding boxes and clothing landmarks for an in-depth analysis framework. A notable feature of DeepFashion is its provision of over 300,000 cross-pose and cross-domain image pairs, offering a unique avenue for nuanced analysis and comparison across different poses and domains, significantly bolstering the scope and depth of the fashion recommendation system being developed in this project.

In this project, the dataset is divided by gender into two distinct categories: male and female. This separation aligns with the project's objective to extract and analyze body proportions, necessitating images that display the full body from head to toe. Additionally, the critical task of skin tone extraction requires the presence of a faces in the images. Consequently, any image lacking a visible face or not showcasing the full body has been excluded from the dataset. Following this selection criterion, the final dataset comprises 770 images in the male category and 3,185 in the female category.

## Feature Extraction

This section delineates the methodologies employed to extract two pivotal features: body proportion and skin tone. It details the systematic procedures and algorithms utilized to extract these attributes from the image dataset, which are fundamental for the subsequent stages of the recommendation system.

### Body Proportion Extraction

MoveNet, a cutting-edge pose estimation model introduced by Google on May 17, 2021, is adept at identifying the spatial positions of specific body parts or keypoints from images or videos, recognizing 17 keypoints like the nose, eyes, and shoulders, among others (Ronny & Na, 2021). It comes in two distinct variants: Lightning and Thunder, each crafted for different scenarios. MoveNet Lightning is optimal for latency-sensitive applications needing real-time performance, particularly when the accuracy demands are moderate whereas MoveNet Thunder is apt for situations requiring accurate pose estimation like in healthcare or professional athletic training where precision is crucial.

Within the scope of this project, the MoveNet Thunder model has been chosen, prioritizing precision over real-time detection capabilities. The decision is informed by the project's exclusive focus on still images rather than video sequences, thereby diminishing the necessity for instantaneous processing. Initially, MoveNet Thunder is deployed to meticulously extract the keypoints from human figures within the images, establishing the foundation for subsequent analyses.

The procedure commences with the identification of all potential keypoints on the human figure. Subsequently, coordinates for crucial keypoints are extracted, namely the left shoulder, right shoulder, left hip, right hip, left ankle, and right ankle. The midpoint between the shoulders is termed as the 'top of torso’ and is derived by averaging the coordinates of the left and right shoulders. This applies to the hip and ankle, termed ‘mid of hip’ and ‘mid of ankle’.

However, there are few instances where not all keypoints are precisely detected, leading to discrepancies in their exact locations or coordinates. Consequently, the derived body proportions may not reflect true measurements, which can affect the reliability of the feature extraction process, and, by extension, the robustness of the recommendations provided by the system. Hence, the affected images are then eliminated from the dataset.

Following this, the shoulder and hip widths are determined by computing the Euclidean distance between the left and right shoulders, and the left and right hips respectively.

The body height is ascertained by summing up two distances: from the top of the torso to the mid of hip, and from the mid of hip to mid of ankle. Leg height is gauged by measuring the Euclidean distance from the mid of hip to the mid of ankle, while torso height is measured from the top of the torso to the mid of hip using the same distance metric.

With these measurements at hand, the vertical body proportion is calculated as the ratio of torso height to leg height. Similarly, the horizontal body proportion is derived from the ratio of shoulder width to hip width. The leg to body ratio is deduced by dividing the leg height by the body height.

In short, the function encapsulates the computed vertical body proportion, horizontal body proportion, and leg to body ratio within a tuple, which it then returns as the output.

A person standing with lines and points

Description automatically generated

Figure 7. An illustration of the extracted key points used in calculating the body proportions.

A computer screen shot of a program code

Description automatically generated

Figure 8. Code snippet of the function “calculate\_propootion”, used in calculating the body proportion.

### Skin Tone Extraction

The procedure commences with the identification of the face within the image, employing the MediaPipe library for accurate facial detection. Upon successful detection of a face, the image is cropped to retain only the facial region, effectively isolating it from the rest of the image.

Subsequent to this, a process of feature refinement is carried out to eliminate irrelevant features that do not correspond to skin. Specifically, regions corresponding to the eyes and mouth are meticulously removed to avoid any potential interference with the accurate determination of skin tone. These removed features are replaced with black pixels, effectively neutralizing their impact on the subsequent analysis of skin color.

Due to the varying quality of images, particularly those of lower resolution, facial detection can be challenging when working with full-body images. To circumvent this challenge, a consistent cropping strategy is employed, targeting the upper half of the images where the face is predominantly located. This approach has significantly improved the detection rate by 35%, as the focus is narrowed to the area of interest. Despite these enhancements, there are instances where facial feature identification is flawed, leading to the exclusion of essential parts of the face, which can affect the accuracy of skin tone data. To address this, manual verification is incorporated into the workflow to correct any misidentified features, thereby ensuring the integrity of the skin tone extraction process. Continuous efforts are made to upgrade the facial detection algorithms, adapting to a variety of image conditions to maintain high accuracy in skin tone extraction across the dataset.

Following the removal of these extraneous features, the color space of the image is converted to HSV (Hue, Saturation, and Value). This conversion is instrumental as the HSV color space is more adept at separating chromatic content (color) from intensity (lightness), which facilitates a more accurate analysis and averaging of skin tones.

In the final step of the procedure, the colour values within the defined HSV (Hue, Saturation, and Value) space are averaged to derive a singular representative value for skin tone. It's crucial to note that the black pixels, which were used to replace the removed features, are excluded from this averaging process to ensure that the calculated skin tone is accurate and representative of the actual skin colour.

A person with a mask and a person with a face mask

Description automatically generated with medium confidence

Figure 9. This illustration outlines the skin tone extraction method: starting with facial detection, it proceeds by discarding non-skin features, converting the facial region to the HSV color model, and finally calculating the average color to determine the individual's skin tone.

A screen shot of a computer code

Description automatically generated

Figure 10. Code snippet of the function extract\_skin\_tone, employed to obtain the skin tone.

The aforementioned feature extraction process is replicated for every image within the dataset. Following the extraction, the collected features are carefully documented and stored within a CSV (Comma-Separated Values) file. These features serve as the foundational elements for the subsequent phase of similarity-based recommendations utilizing the K-Nearest Neighbors (KNN) algorithm and autoencoders.

## Algorithm Used for Recommendation

This section is dedicated to a detailed exploration of the algorithms that form the backbone of the recommendation process. It delves into methods and strategies employed to provide personalized fashion suggestions, highlighting how these algorithms leverage the extracted features to generate precise and user-centric recommendations.

### K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm serves as the cornerstone for the recommendation system in this project. With the objective of identifying the top five images that exhibit a high degree of similarity to the input image in terms of body proportion and skin tone, KNN emerges as a fitting choice given its efficacy in handling multi-feature similarity computations.

The KNN algorithm is strategically deployed to perform similarity searches rather than clustering. By calculating distances based solely on skin tone and body proportion features, the system efficiently pinpoints and retrieves fashion items that closely match the user's physical attributes. When a user interacts with the system, it doesn't group items into clusters; instead, it functions like a precise compass, directing the user to those items that lie nearest in the feature space defined by skin tone and body proportions. This focused approach ensures that each recommendation is a reflection of the user's unique measurements and coloring, offering a personalized selection that is likely to resonate with their individual aesthetic and fit requirements.

Prior to training the KNN model, an essential preprocessing step is undertaken to normalize the dataset. This process adjusts the features to a common scale, preventing any single feature from dominating the distance calculations due to its scale. Normalization is particularly important in this context where the features vary widely in magnitude. By scaling these features into a uniform range, typically [0, 1], the algorithm can evaluate the similarity between images in a balanced manner, ensuring that each feature contributes equally to the overall distance computation.

Initially, the dataset, comprising extracted features such as hue, saturation, value (representing skin tone), and three distinct body proportion metrics, is loaded from a CSV file. These features collectively constitute the multi-dimensional space within which the similarity comparison takes place. The KNN model utilizes this dataset to determine the proximity between images. When an input feature is provided, the KNN algorithm searches through the dataset to find the five nearest neighbors based on Euclidean distance in this six-dimensional feature space, effectively capturing the similarities of color and proportion.

By employing the KNN algorithm, the system is positioned to efficiently traverse through the feature space and search the images that align closely with the user's physical attributes as captured in the input image. This methodical approach facilitates a more personalized and precise recommendation, enhancing the user's journey in discovering fashion items that resonate with their unique aesthetic and physique.

A computer code with green and yellow text

Description automatically generated

Figure 11. Code snippet of how KNN is used to find the 5 nearest neighbours.

In addition to Euclidean distance, the system employs a diverse array of distance metrics to analyze similarity, encompassing Manhattan and Minkowski distance. Manhattan distance, also known as city block distance, aggregates the absolute differences between coordinates. Minkowski distance generalizes different distance metrics through an adjustable parameter; it can represent both Euclidean distance when the parameter is set to two and Manhattan distance when set to one. This multifaceted approach to similarity measurement allows for a comprehensive assessment of the relationships between data points.

### Autoencoder

In parallel to the KNN algorithm, the project leverages an autoencoder neural network to distill the essence of the dataset into a more compact, yet informative representation. The autoencoder consists of two principal components: an encoder, which compresses the input data into a lower-dimensional latent space, and a decoder, which attempts to reconstruct the input data from this compressed form. The merit of employing an autoencoder lies in its ability to learn a dense encoding of the data, capturing intrinsic patterns that may not be apparent when considering the raw features directly.

The encoder part of the autoencoder transforms the high-dimensional data into a lower-dimensional space, which not only reduces the computational load but also can help in mitigating the curse of dimensionality that often hampers KNN's performance. The encoded features represent a distilled version of the data where the most significant patterns are preserved.

The autoencoder is trained through a process of optimization where the loss function aims to minimize the reconstruction error, ensuring that the encoded representation retains as much relevant information as possible. Once trained, the encoder provides a new set of features for the input data, which can be used to compute similarities. Distances between the encoded representation of the input image and those of the images in the dataset are calculated, typically using Euclidean metrics, to find the closest matches. The underlying architecture of the encoder is demonstrated through the code snippets presented below, which detail the specific structural components and configurations.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer code

Description automatically generated

Figure 12. This figure provides snippets of code illustrating the definition and training phases of an autoencoder. It highlights the setup of the neural network layers, parameter specification, and the execution of the training routine.

At the core of the encoder is an input layer designed to accept feature vectors of size six, corresponding to the number of attributes (hue, saturation, value, and three body proportions) extracted from each image. The input layer serves as the gateway through which data enters the network.

Following the input layer is a dense layer, which is a fully connected neural network layer that comprises 32 neurons. This layer is responsible for transforming the input data into an encoded representation. The activation function for this dense layer is the Rectified Linear Unit (ReLU), which introduces non-linearity into the network, allowing it to learn complex patterns within the data. ReLU was chosen for its efficiency and effectiveness in many neural network applications.

This dense layer effectively serves as the encoder, compressing the six-dimensional input data into a 32-dimensional encoded vector. The encoded vector is a more compact representation of the input data, capturing the most important features necessary for reconstructing the input with minimal loss of information.

The subsequent structure of the autoencoder is the decoder, which mirrors the encoder in reverse. It consists of another dense layer with the same number of neurons as the input layer, tasked with reconstructing the input data from the encoded representation. The activation function for this layer is the sigmoid function, which ensures that the output of the decoder lies between 0 and 1, matching the normalized range of the input data.

The autoencoder is compiled with the Adam optimizer, a widely used algorithm for training neural networks due to its adaptive learning rate capabilities. The loss function employed is the mean squared error (MSE), which quantifies the difference between the input vectors and their reconstructed counterparts from the decoder. The use of MSE is particularly apt for regression tasks where the goal is to minimize the error in the reconstruction.

To prevent overfitting and to ensure that the model generalizes well, an early stopping callback is defined. This mechanism monitors the validation loss and halts the training process if no improvement is observed after a predefined number of epochs, in this case, ten. It also restores the weights from the epoch that achieved the best validation loss, ensuring the model retains the most effective learned parameters.

In summary, the encoder is a critical component of the autoencoder architecture, tasked with compressing high-dimensional data into a lower-dimensional, encoded form that retains significant information for accurate reconstruction. Its design and training are pivotal for the success of the autoencoder in learning efficient representations of the dataset, which in turn facilitates more refined similarity measurements for the recommendation system.

After the autoencoder is trained, the input image is normalized and encoded to align with the training data's scale. The encoded image is then compared with the encoded dataset using a variety of distance metrics such as Euclidean, Manhattan, and Minkowski. The five images that demonstrate the smallest distance across these metrics are identified as the closest matches. This comparison process and the resulting selections are illustrated in the figure that follows, offering a visual representation of the methodological steps involved.

A screenshot of a computer program

Description automatically generated

Figure 13. This figure displays code snippets from the 'get\_recommendations' function, showcasing the retrieval of the top 5 images most similar to the input image within the encoded feature space, as determined by the trained autoencoder model.

By incorporating the autoencoder's perspective, the recommendation system gains an additional layer of abstraction. It can discern subtle similarities between images that might be overlooked by KNN when operating solely in the original feature space. This dual approach, employing both KNN and the autoencoder, enriches the recommendation process, offering users a set of images that are not only similar in the conventional sense but also akin in the nuanced, learned representation of the autoencoder.

# Results and Discussion

This section presents a comprehensive analysis of the outcomes derived from the implemented methodology, offering a critical evaluation and discussion of the findings. It aims to interpret the significance of the results in the context of the project's objectives, discussing the effectiveness of the feature extraction and machine learning techniques used, as well as the performance of the recommendation system developed.

## K-Nearest Neighbours

The input image selected for testing originates from the identical dataset employed to train the k-Nearest Neighbors algorithm. The model/human depicted in the input image is featured in various other images within the dataset, each showcasing different fashion ensembles. Ideally, the system is designed to recommend images showcasing the same model adorned in a variety of outfits.

To visualize the position of input features within the multidimensional feature space, a scatter plot can be employed. However, the inherent complexity of six-dimensional data necessitates the use of dimensionality reduction techniques to enable such a visualization. The t-Distributed Stochastic Neighbor Embedding (t-SNE) is particularly adept at condensing high-dimensional data into two dimensions while preserving the relative distances between points. By applying t-SNE, we can project the six features onto a two-dimensional plane. The subsequent figure provides a graphical representation of this transformation, illustrating the location of each data point in relation to the input features within the newly defined two-dimensional feature space.

A blue and white background with black text

Description automatically generated with medium confidence

Figure 14. This figure demonstrates the dimensionality reduction from a 6-dimensional feature space to a 2-dimensional plane using t-SNE, where the red point signifies the input data, and the blue points represent the dataset.

Subsequent to this, the figures presented below illustrate the outcomes of applying various distance metrics to identify the five most similar images, as determined by their respective distances within the feature space.

**Euclidean Distance**

A group of men standing in different poses

Description automatically generated

Figure 15. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Euclidean distance metric in the feature space.

The initial recommendation perfectly mirrors the input image, leading to a zero-distance measure as anticipated. Subsequent recommendations, specifically the 2nd, 4th, and 5th images, display the model in alternate attire, affirming the efficacy of the recommendation system. These results underscore the system's capability to discern and associate images of the same model, thereby validating its practical utility in delivering relevant fashion recommendations.

**Manhattan Distance**

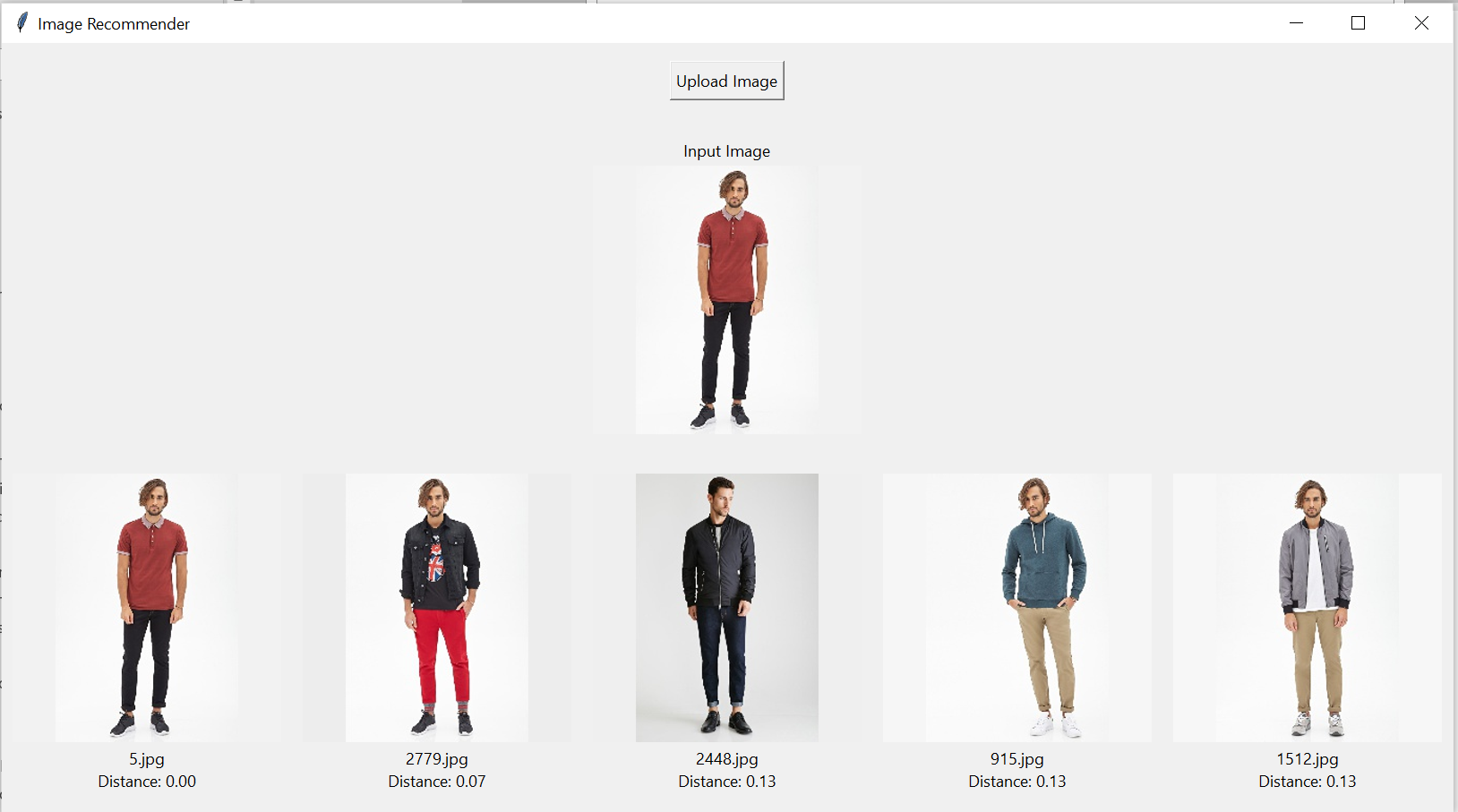


Figure 16. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Manhattan distance metric in the feature space.

The recommendations generated using the Manhattan distance metric exhibit subtle variations compared to those obtained through Euclidean distance. Notably, the image previously ranked fourth in the Euclidean scheme has been relegated to the fifth position within the Manhattan framework. Meanwhile, the fourth spot is occupied by a different image that also features the same model. This shift suggests that the Manhattan distance, with its emphasis on the sum of absolute differences, offers a slightly altered perspective on similarity that may be sensitive to certain feature disparities.

**Minkowski Distance**

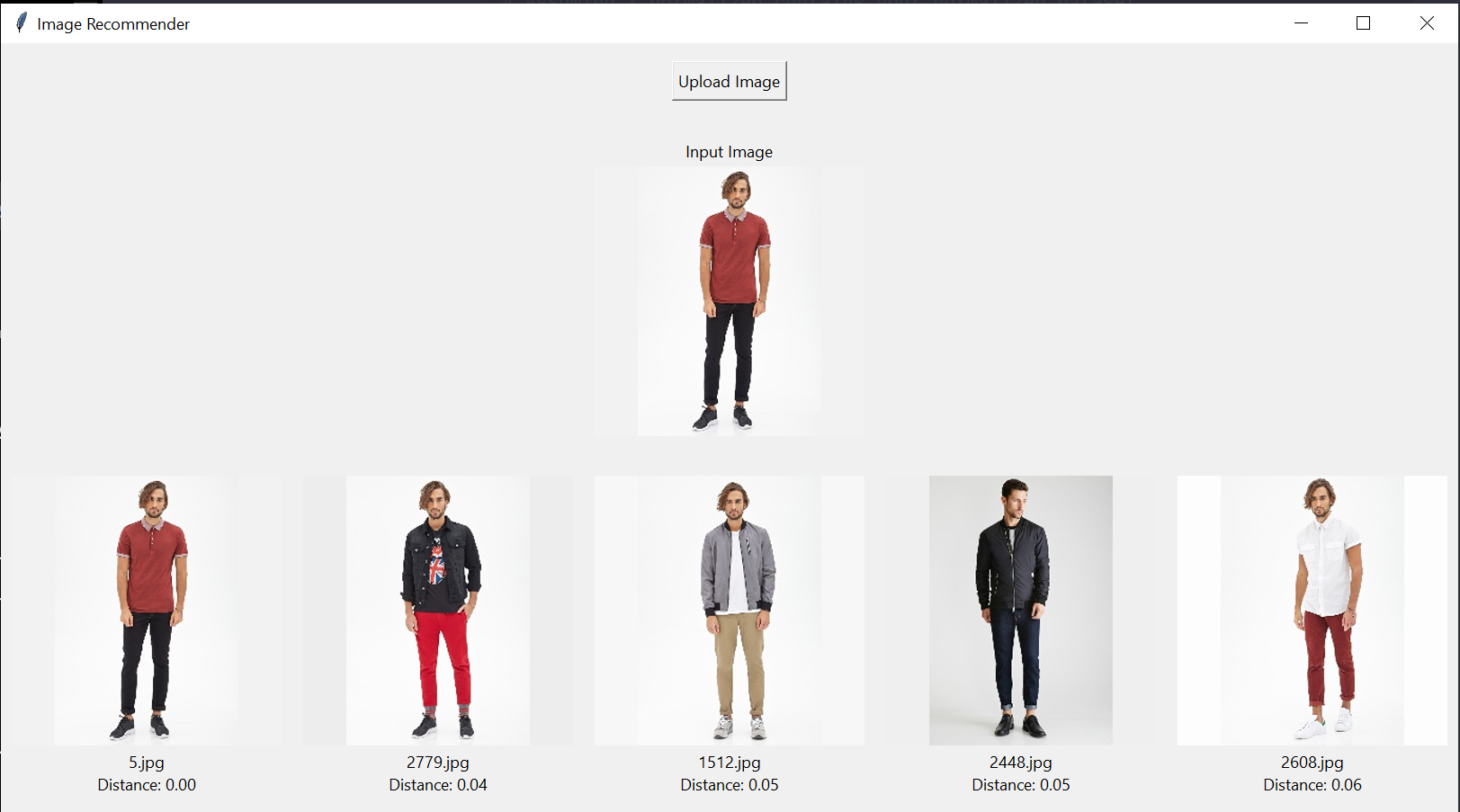


Figure 17. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Minkowski distance metric in the feature space.

Using the Minkowski distance metric, the system effectively identifies different images of the same model wearing various outfits, as seen in the second, third, and fifth recommendations. This observation suggests that the Minkowski distance, with its flexibility to adapt between Euclidean and Manhattan distances, is a strong candidate for assessing similarity in fashion images. Its effectiveness in highlighting the same model in diverse attire underscores its practicality for fashion recommendations, demonstrating its capacity to handle the subtle complexities of such a visual dataset.

In summary, these nuances between the three metrics highlight the importance of selecting an appropriate distance measure that aligns with the specific nuances and requirements of the recommendation task at hand. It is vital to acknowledge that these distances cannot be directly compared, as each metric operates on a distinct scale and interprets the feature space differently. Thus, these comparisons serve more as a reference to visualize the impact of varying metrics on the outcomes of image retrieval. This insight can be leveraged to fine-tune the recommendation system, potentially by incorporating a weighted combination of different distance metrics to achieve a more balanced and representative model of similarity.

## Autoencoder

The training progress of the autoencoder is illustrated by a loss graph, which plots the model's learning curve across epochs. Each epoch corresponds to a full cycle of learning, during which the model refines its internal parameters to reduce the loss—a measure of the difference between the original and reconstructed data. The graph reveals a steady diminishment of both training and validation loss, from 0.1186 to 0.0015 and 0.1151 to 0.0010, respectively, across all 300 epochs. This consistent decline in loss signifies the autoencoder's capability to more accurately compress and decode the input data. Notably, the concurrent reduction of both training and validation losses suggests effective model generalization. In other words, the model is learning patterns representative of the dataset at large, rather than memorizing the training inputs, thereby promising a robust performance on new, unseen data.

A graph of a line

Description automatically generated

Figure 18. Graph of training and validation loss against epoch throughout the training process of the autoencoder.

The image employed as input in the kNN algorithm serves a similar role within the autoencoder framework. Initially, the feature set of the reference image undergoes normalization and subsequent encoding. This encoded representation is then leveraged to compute the proximity to other images within the encoded feature space.

In the context of the autoencoder framework, the input image undergoes a transformation similar to the kNN approach. The feature set extracted from the image is first normalized, aligning it with the scale of the dataset, and then encoded into a 32-dimensional latent space. This high-dimensional encoded representation, while rich in information, exceeds our visual representation capabilities. To address this, we employ t-SNE for dimensionality reduction, facilitating a two-dimensional visualization that preserves the relative distances within the original 32-dimensional space. This allows us to plot not only the entirety of the dataset but also to distinctly highlight the position of the input image among the data points, providing an intuitive graphical representation of its proximity to other images in the feature space.

A screen shot of a diagram

Description automatically generated

This figure demonstrates the dimensionality reduction from a 32-dimensional feature space to a 2-dimensional plane using t-SNE, where the red point signifies the input data, and the blue points represent the dataset.

The process continues in the identification of the five most analogous images, each determined by distinct distance metrics. The outcomes of these computations, using varying metrics, are systematically illustrated in the subsequent figures.

**Euclidean Distance**

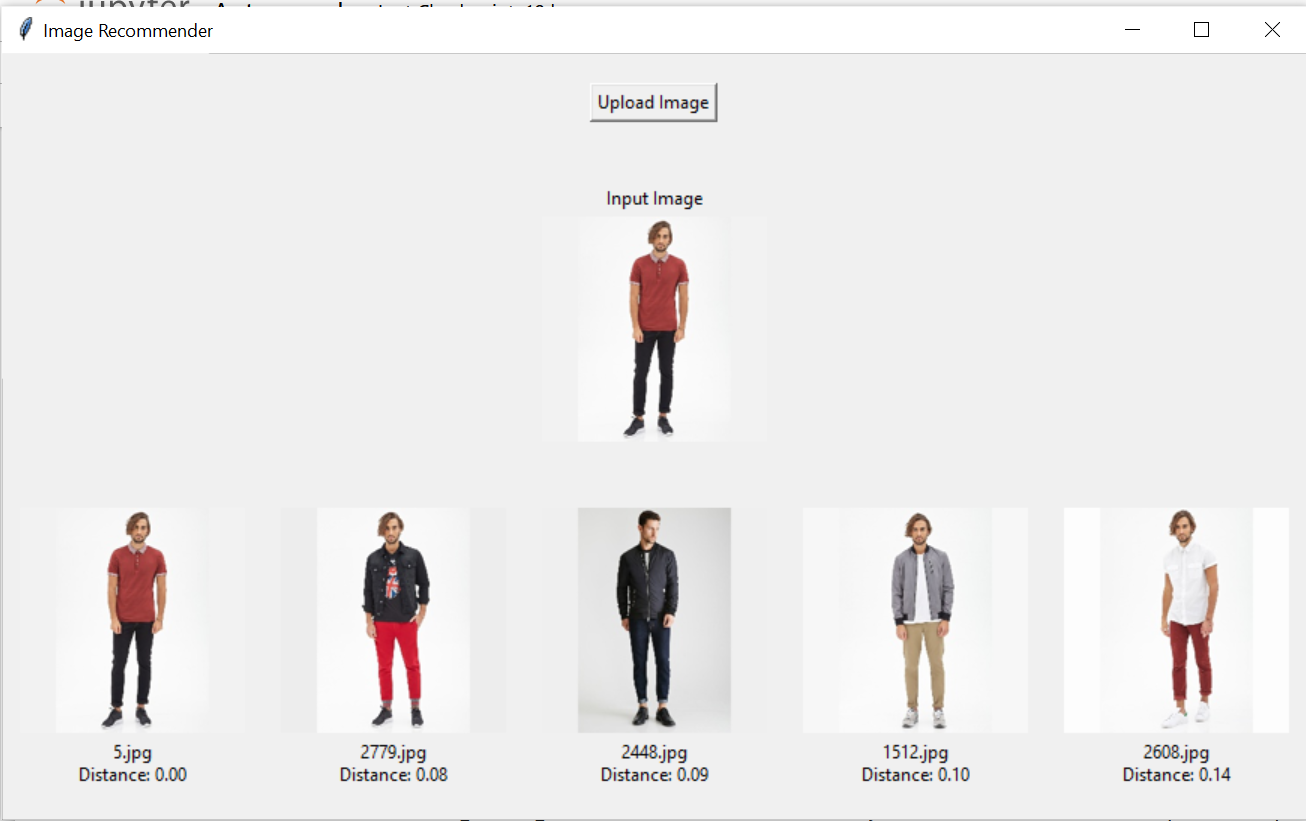


Figure 19. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Euclidean distance metric in the encoed feature space.

The recommendations generated using Euclidean distance in the autoencoder model are consistent with those obtained from the kNN algorithm, which also utilizes Euclidean distance. This coherence reinforces the reliability of the Euclidean metric in maintaining similarity across both methodologies.

**Manhattan Distance**

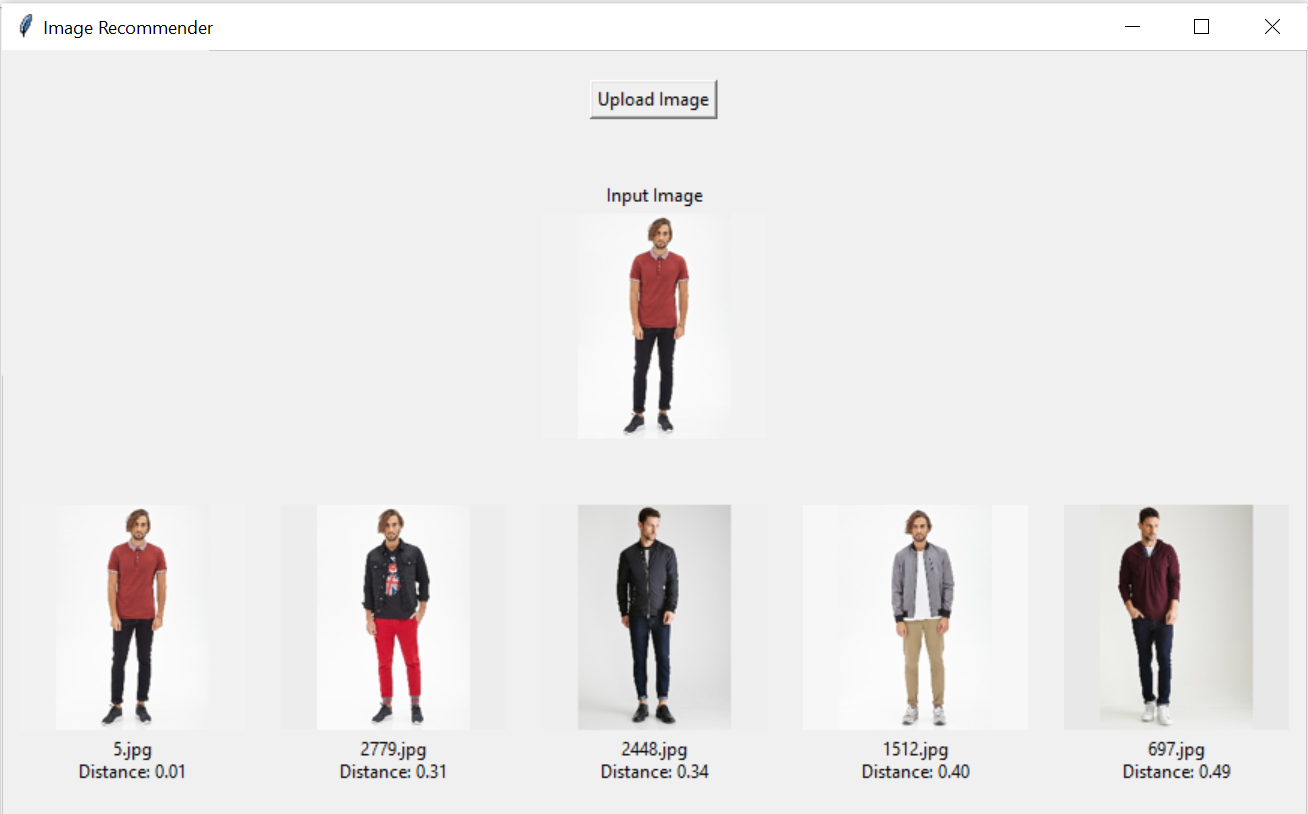


Figure 20. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Manhattan distance metric in the encoed feature space.

The recommendations generated using the Manhattan distance metric display a variance from those produced with Euclidean distance. Aside from the identical first image, the second and fourth images in the Manhattan-based recommendations correctly feature the same model as in the input image. This discrepancy suggests that while Manhattan distance provides useful results, it may yield recommendations with a margin of error when compared to those obtained through Euclidean distance. This highlights the importance of metric selection in the accuracy of image retrieval systems.

**Mitkowski Distance**

A screenshot of a computer screen

Description automatically generated

Figure 21. This figure showcases the top 5 images that are most similar to a reference image, selected based on the Mitkowski distance metric in the encoed feature space.

When employing the Minkowski distance metric, a similar pattern emerges as observed with Manhattan distance. The second and fourth images within the Minkowski-based recommendations accurately depict the same model present in the input image. Such observations are crucial in understanding the comparative strengths of various distance metrics in the context of image retrieval tasks.

## Evaluation of Different Classes of Skin Tone

In order to assess the system's adaptability and accuracy across a spectrum of skin tones, test images featuring models with diverse skin tones were utilized. The results delineated in the subsequent figures, provide a visual representation of the system's performance in recognizing and accurately responding to variations in skin tone. This evaluation is pivotal for ensuring that the system's capabilities are not biased towards any specific skin tone, thus affirming its reliability and inclusiveness.

**Dark Skin tone**

A group of men standing in different poses

Description automatically generated

A screenshot of a person

Description automatically generated

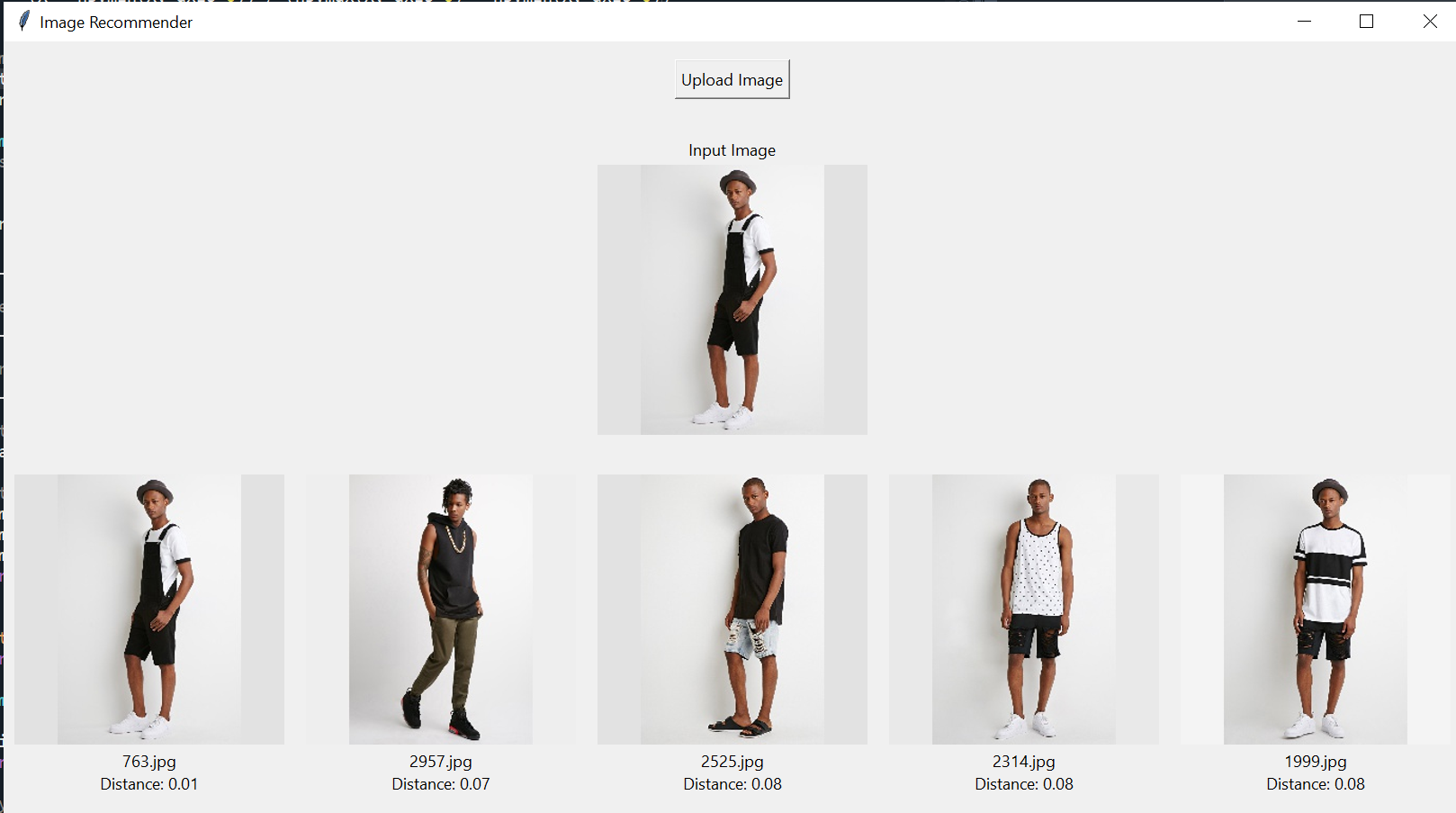


Figure 22. This figure illustrates the system's recommendations of fashion items that complement dark skin tones, highlighting the algorithm's adaptability to skin tone variability.

**Fair Skin tone**

A screenshot of a computer screen

Description automatically generated

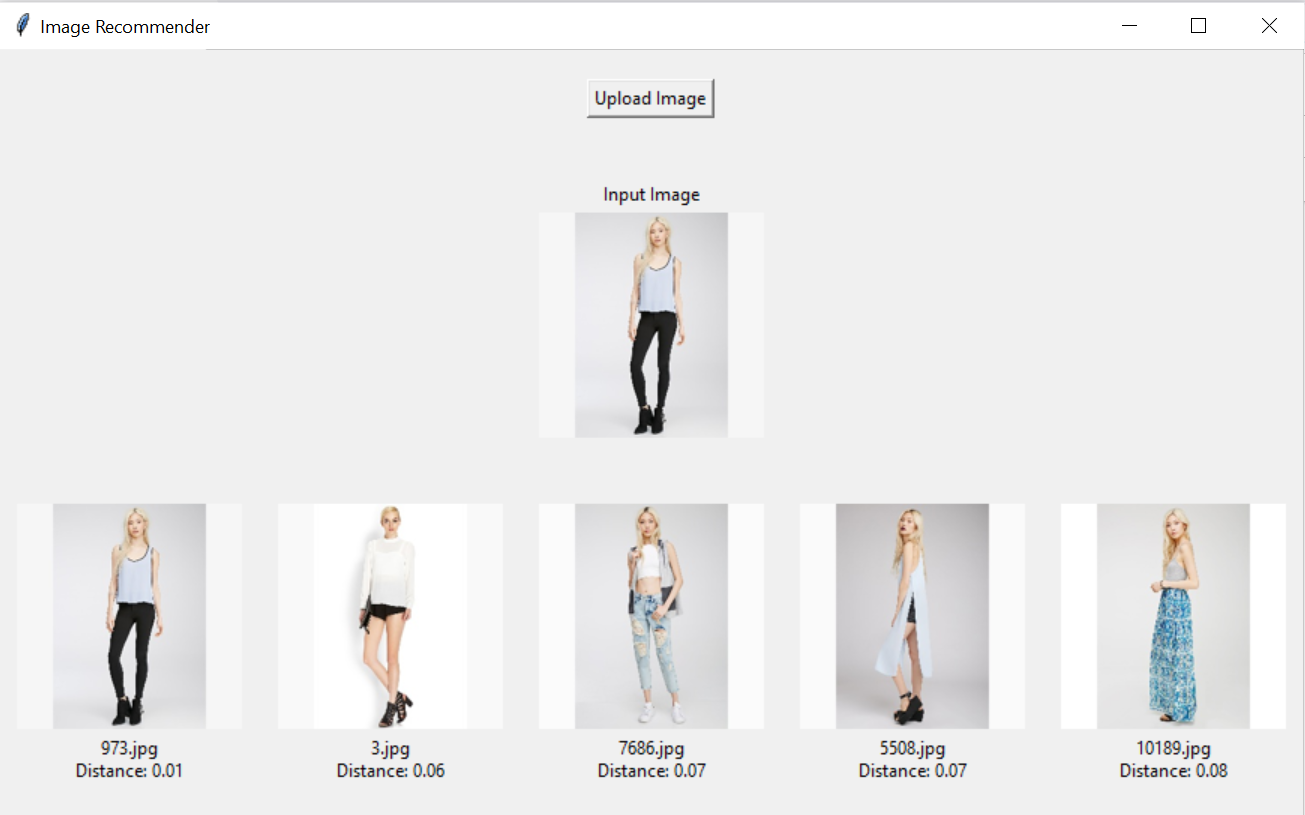
A screenshot of a computer screen

Description automatically generated

Figure 23. This figure illustrates the system's recommendations of fashion items that complement fair skin tones, highlighting the algorithm's adaptability to skin tone variability.

## Evaluation Across Genders

The system's versatility is further demonstrated by its performance in processing images of women. This ensures a comprehensive understanding of the system's capability to handle gender diversity effectively. By extending the evaluation to include various genders, the system's unbiased performance and adaptability can be critically assessed, ensuring its applicability to a broader demographic.



A screenshot of a person

Description automatically generated

Figure 24. This figure illustrates the system's recommendations of fashion items that complement female, highlighting the algorithm's adaptability to gender variability.

## Evaluation with Daily Life Images

To validate the system's adaptability to real-world conditions, an evaluation was conducted using images from daily life. This included images captured in less controlled environments, such as a selfie contributed by the researcher. Such testing is critical to verify the robustness of the system when applied to commonplace scenarios, thereby ensuring its utility extends beyond laboratory settings.

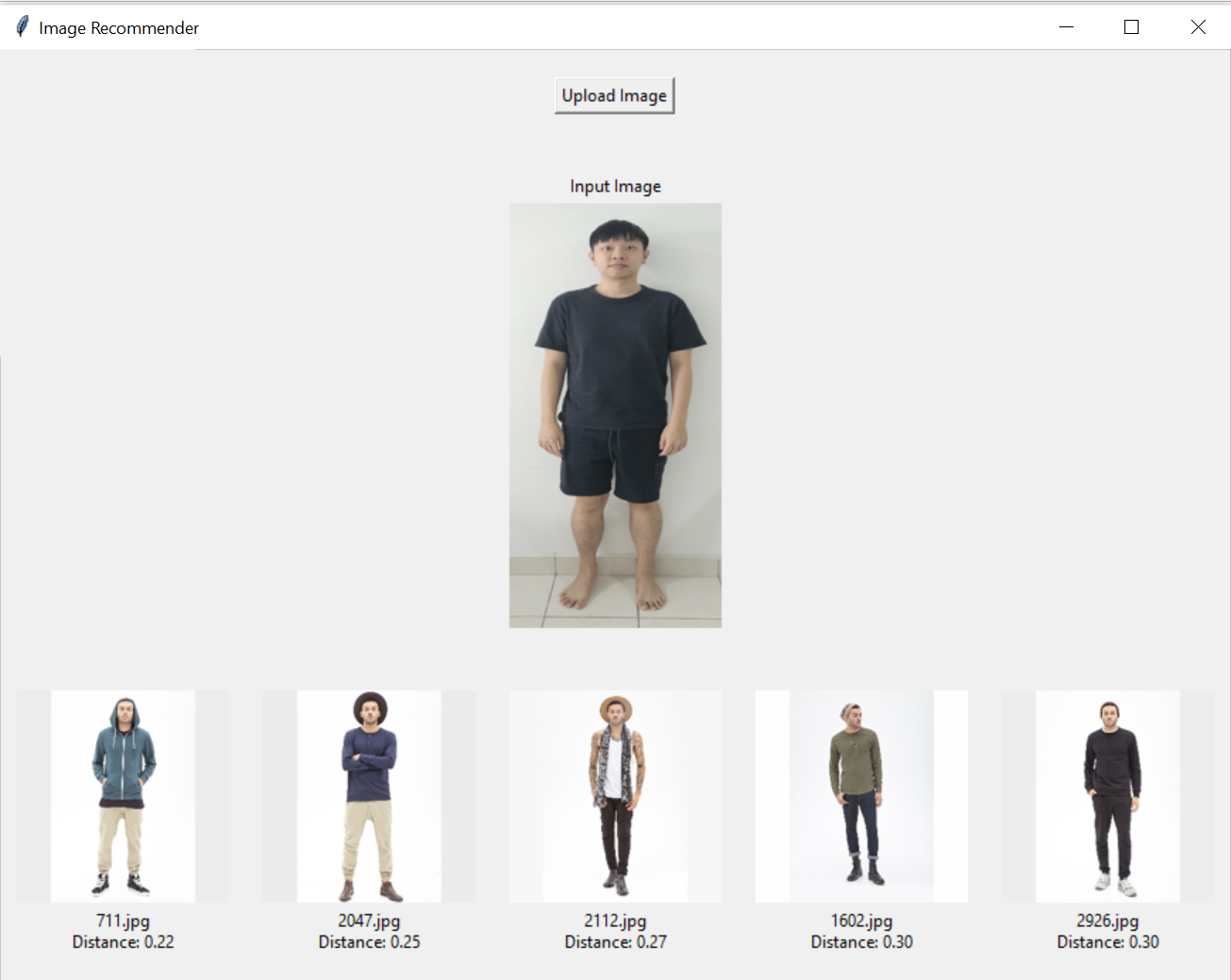


Figure 25. This figure displays the system's capacity to recommend fashion items based on images from daily life, demonstrating the robustness and practical application of the recommendation algorithm.

In summary, the devised system exhibits robust adaptability across various dimensions, including gender, skin tone, and body proportions. However, it is essential to note that the recommendations derived from the autoencoder and kNN are not directly comparable due to the distinct methodologies they employ. Future assessments of the system's efficacy could be enriched by incorporating user feedback, which would provide valuable insights into the user-centric performance of the recommendation engine. Further explorations may involve investigating the system's scalability and its performance with an expanded dataset, or examining the integration of additional features that could refine the recommendations, such as contextual information or personalization aspects based on user preferences and interaction history.

## Comparison with Related Works

In contrast to the approach by (Atharv , Kunal , Manav , & Neha , 2020)which necessitates manual input of skin tone by users for personalized recommendations, the system developed in this project boasts an automated skin tone extraction capability. This advancement allows for a more streamlined and user-friendly experience, as it eliminates the need for manual classification, enhancing the ease of use and reducing potential biases associated with self-reported data.

Furthermore, the research by (Digant, Anushree , Monali , Shivani , & Tabassum , 2019) and (Reeta, Anisha , Tejashri , & Siddesh , 2021) partitions skin tones into five dominant categories, from which garment recommendations are made to match these broad classifications. In contrast, the current proposed system bypasses broad categorization and instead, directly extract and comparing skin tones to find the closest image matches. This method provides a more tailored recommendation, potentially improving the accuracy and personal relevance of the results.

Similarly, while the study by Seema et al. involves categorizing users into one of eight body shapes to recommend appropriate clothing, the proposed system here does not confine users to predefined body shape categories. It instead focuses on the precise extraction of body proportions from user images, aiming to locate and suggest garments that share the closest resemblance in proportions. This not only affords a higher level of specificity in recommendations but may also contribute to a more inclusive system that can cater to a wider range of body types, thereby enhancing the personalized shopping experience for each user.

Moreover, the system designed in this project is characterized by its adaptability, capable of integrating new fashion trends by continuously updating its image database. This ensures that recommendations remain current and in line with the latest styles. Furthermore, the system's strength lies in its comprehensive feature extraction process. By employing encoder methodologies for dimensionality reduction, it captures the intricate details of user-uploaded images, which are essential for crafting accurate and personalized recommendations. This stands in notable contrast to systems that rely solely on CNNs or RNNs for feature extraction (Atharv , Kunal , Manav , & Neha , 2020). The encoder approach can discern subtle nuances in the images, which may include textures, patterns, and the interplay of colors and shapes, thereby providing a sophisticated level of analysis that can significantly enhance the personalization of the recommendation system. This capability underlines the system's versatility and potential to offer a more refined and individualized user experience.

## Limitation

The system's efficacy is significantly contingent upon the comprehensiveness and diversity of the dataset employed. Currently, the recommendations are confined to matching with pre-existing images within the dataset, which may not encompass the full spectrum of body types, skin tones, or styles. This limitation potentially leads to recommendations that may not be universally suitable or inclusive, particularly if the dataset lacks variability or size. Moreover, the system does not inherently account for the context of garment usage, such as the occasion or weather conditions, unless such factors are explicitly encoded within the dataset. This absence of contextual awareness could result in recommendations that, while visually similar to the input image, may not align with the user's situational needs.

Another limitation is the subjective nature of clothing suitability. The system assumes that garments that appear suitable for the models in the dataset will be equally suitable for the user. This assumption overlooks individual preferences and the nuances of personal style, potentially limiting the system's ability to provide truly personalized recommendations.

In terms of feature extraction, the process could be enhanced by incorporating neural network models to segment the human figure prior to feature extraction. This refinement could lead to more focused and accurate recognition of key attributes, such as facial features and body proportions, thereby enhancing the personalization of recommendations.

The system also lacks a dynamic learning component that adjusts to changes in user preferences over time. Without a feedback mechanism to learn from user interactions and adapt to evolving tastes, the system may not fully capture the individual's style evolution.

Lastly, the challenge of objectively evaluating the system's performance remains. In the domain of fashion, where preferences are highly subjective, establishing quantitative metrics for recommendation effectiveness is complex. Future work could aim to develop more evaluation criteria that consider user satisfaction and the relevance of recommendations in real-world scenarios. These limitations highlight critical areas for potential refinement and underscore the importance of ongoing development to ensure the system's adaptability and user-centric performance.

# Conclusion and Future Works

This final chapter encapsulates the essence and the findings of the research, summarizing the significant contributions this study makes to the field of personalized fashion recommendation systems. It revisits the research objectives outlined in the introduction, reflects on the methodologies employed, and assesses the success of the project in addressing the problem statement. By synthesizing the insights obtained from the research, this chapter provides a conclusive statement on the implications of integrating user's skin tone and body proportions into a fashion recommending algorithm. Furthermore, it paves the way for future research that can build upon the groundwork laid by this thesis, aiming to refine and enhance the capabilities of AI-driven personalization in fashion technology.

## Conclusion

This project has successfully addressed the challenge of providing personalized fashion recommendations through the development of a sophisticated machine learning and deep learning-based system. The primary aim to enhance personalization in the fashion industry by utilizing users' skin tones and body proportions has been achieved through a multi-faceted approach. The following objectives have been met, supporting the aim of the study:

* **Objective #1**: Segment and extract the face portion from the images to determine skin tone by averaging the color values.

The system effectively segments and extracts the facial portion from user images to determine skin tone. The average color values are computed, thereby allowing for an accurate representation of the user's skin tone without the need for manual input, as seen in previous studies.

* **Objective #2**: Utilize pose estimation techniques to accurately extract body proportions from user-supplied images.

Through the application of pose estimation techniques, the system is capable of accurately extracting body proportions from the images provided by the users. This facilitates a more nuanced approach to personalization, moving beyond the standard categorization into predefined classes of body shapes.

* **Objective #3**: Provide fashion recommendations that complement these physical attributes using Content-Based Filtering.

Employing Content-Based Filtering, the system recommends fashion items that align with the extracted physical attributes. This method contrasts with previous works that required manual classification of skin tones and body shapes, offering a dynamic and automated approach to fashion recommendations.

The system's versatility is showcased in its ability to integrate new trends by simply adding new images, thus remaining current and relevant in a fast-paced industry. The utilization of encoders for dimensionality reduction captures the intricacies of the user images more effectively compared to systems that solely rely on traditional CNNs or RNNs. The system, through Content-Based Filtering, offers a user-centric recommendation experience that surpasses the capabilities of existing manual and semi-automated systems. It demonstrates the ability to provide accurate and personalized fashion advice without the constraints of static datasets.

In short, the deployment of these algorithms within the project underscores the potential and effectiveness of AI and machine learning techniques in customizing fashion recommendations. This system not only advances the state of the art but also provides a scalable and adaptable framework for future enhancements in the realm of fashion technology.

## Future Works

The future work for the project is multifaceted and aims at enhancing the capabilities of the system and addressing its current limitations. An immediate area for advancement involves the expansion and diversification of the dataset. A broader dataset would enable the system to offer a more inclusive array of fashion recommendations, accommodating a vast spectrum of body types, skin tones, and fashion trends. This enlargement should also consider the dynamic nature of fashion by integrating real-time trend data, thus keeping the recommendations fresh and relevant.

Another significant upgrade is relevant to the system's feature extraction capabilities. The integration of more advanced neural network models could refine the segmentation process, focusing on detailed attributes of the user's physique and the intricacies of clothing patterns. Such advancements would likely yield a marked improvement in the accuracy of personalized recommendations.

The incorporation of a user feedback mechanism stands as a critical enhancement that would allow the system to learn from the users' responses to its recommendations. By adapting to users' preferences over time, the system could offer increasingly relevant suggestions, thereby enhancing the user experience and the system’s learning algorithms.

Further, to assess the system's applicability in real-world scenarios, comprehensive user testing is essential. Such testing would gauge the system's performance across diverse environments and user groups, providing invaluable data to refine the recommendation algorithms based on actual user interactions and satisfaction.

Finally, the exploration of virtual and augmented reality technologies could revolutionize the user experience by offering immersive and interactive visualization capabilities. Such features could include virtual try-ons, allowing users to preview recommended fashion items in a virtual space, thereby enhancing decision-making and user engagement.

In conclusion, the proposed future enhancements would not only address the current limitations but also significantly advance the system's technological edge. By continually refining the system through user feedback, expanding its reach, and incorporating cutting-edge technologies, the project has the potential to set a new benchmark in personalized fashion recommendation systems.

# References

Aleksandr , P., & Craig , M. (2022). A Systematic Review and Replicability Study of BERT4Rec for Sequential Recommendation. *Sixteenth ACM Conference on Recommender Systems (RecSys ’22*. doi: https://doi.org/10.1145/3523227.3548487.

Alhamza , A., Rozmie , R., Mosleh , A., & Mohammed , A. (2016). A Preliminary Performance Evaluation of K-means, KNN and EM Unsupervised Machine Learning Methods for Network Flow Classification. *International Journal of Electrical and Computer Engineering (IJECE)*, 778-784. doi:10.11591/ijece.v6i2.pp778-784

Aneesh K, P V Rohith , K., Sai , N. U., & Archana, N. (2022). Fashion Recommendation System. *Ijraset Journal For Research in Applied Science and Engineering Technology*. doi:https://doi.org/10.22214/ijraset.2022.444362

Atharv , P., Kunal , G., Manav , J., & Neha , K. (2020). A Review on Clothes Matching and Recommendation Systems based on user Attributes. *International Journal of Engineering Research & Technology (IJERT)*. doi:10.17577/IJERTV9IS080371

Bhure, B. P., Bansod, P. T., Amgaokar, M. S., Lodiwale, S. P., Orkey, A. P., & Mohod, A. (2021). A Review on Outfit Fashion Recommendation System. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 220-222. doi:10.32628/CSEIT217368

Chakraborty, S., Hoque, M., Jeem, N. R., Biswas, M. C., Bardhan, D., & Lobaton, E. (2021). Fashion Recommendation Systems, Models and Methods: A Review. *Informatics*, 8-49. doi:https://doi.org/10.3390/

Cristina , B., Jesús , F.-B., Javier , D., Henry , A., Félix , J., & María, J. (2022). A comparative analysis of pose estimation models as enablers for a smart-mirror physical rehabilitation system. *Procedia Computer Science*, 2536-2545. doi:https://doi.org/10.1016/j.procs.2022.09.312.

Daud , I., Manishankar , P., Sandeep, K., & Bikal , C. (2022). Estimate human body measurement from 2D image using computer vision. *Journal of Emerging Technologies and Innovative Research (JETIR)*.

Deepjyoti , R., & Mala, D. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*. doi:https://doi.org/10.1186/s40537-022-00592-5

Deldjoo, Y., Nazary, F., Arnau Ramisa, McAuley, J., Pellegrini, G., Bellog, A., & T.D.Noia. (2022). A Review of Modern Fashion Recommender Systems. *ArXiv*.

Diana , F., Sofia, S., António , A., & José , M. (2020). Recommendation System Using Autoencoders. *Appl. Sci.* doi:https://doi.org/10.3390/app10165510

Diana, B., Adrian , D., & Radu , D. (2018). Automatic Skin Tone Extraction for Visagism Applications. *n Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018)*, 466-473. doi:10.5220/0006711104660473

Digant, G. P., Anushree , K., Monali , D., Shivani , L., & Tabassum , M. (2019). Skin-tone And Occasion Oriented Outfit Recommendation System. *2nd International Conference on Advances in Science & Technology (ICAST-2019)*. doi:http://ssrn.com/link/2019-ICAST.html

Dor , B., Noam , K., & Raja , G. (2020). Autoencoders. *Computer Vision and Pattern Recognition*. doi:https://doi.org/10.48550/arXiv.2003.05991

Euse, E. (7 June, 2016). *Men Will Spend Four Months of Their Lives Deciding What to Wear*. Retrieved from complex: https://www.complex.com/style/a/erica-euse/men-spend-four-months-of-lives-deciding-what-to-wear

Fei, S., Jun, L., Jian, W., Chang Hua, P., Xiao Lin, Wenweu, O., & Peng, J. (2019). BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. *CIKM '19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1441-1350.

Guan, C., Qin, S., Ling, W., & Long, Y. (2018). Enhancing Apparel Data Based on Fashion Theory for Developing a Novel Apparel Style Recommendation System. *Rocha, Á., Adeli, H., Reis, L., Costanzo, S. (eds) Trends and Advances in Information Systems and Technologies. WorldCIST'18 2018, 747*. doi:https://doi.org/10.1007/978-3-319-77700-9\_4

Haoming , C., Runyang , F., Sifan, W., Hao, X., & Fengcheng, Z. (2022). 2D Human Pose Estimation: A Survey. *arvix: Computer Vision and Pattern Recognition*.

Huang, E. (9 November, 2021). *Building an AI-Powered Outfit Recommendation System With Dataiku*. Retrieved from dataiku: https://blog.dataiku.com/outfit-recommendation-system

Jaechoon , J., Seolhwa , L., Chanhee , L., & Dongyub, L. (2020). Development of Fashion Product Retrieval and Recommendations Model Based on Deep Learning. *Electronics*. doi:https://doi.org/10.3390/electronics9030508

Jinbao , W., Shujie , T., Xiantong , Z., Shuo , X., Feng , Z., Zhenyu , H., & Ling , S. (2021). Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding,Volume 210,*. doi:https://doi.org/10.1016/j.cviu.2021.103225.

Jingwen , S., Weixing , D., & Niancai , S. (2018). A Survey of kNN Algorithm. *Information Engineering and Applied Computing*.

Leonardo , A. A., Mateus , F., Paulo, H., & Wellington , S. M. (2018). A Fast Similarity Search kNN for Textual Datasets. *2018 Symposium on High Performance Computing Systems (WSCAD)*. doi:10.1109/WSCAD.2018.00043

M Sridevi, N ManikyaArun, M Sheshikala, & Sudarshan E. (2020). Personalized fashion recommender system with image based neural networks. *IOP Conf. Series: Materials Science and Engineering 981 (2020) 022073*. doi:10.1088/1757-899X/981/2/022073

M. A. Chyad, H. A. Alsattar, B. B. Zaidan, A. A. Zaidan, & Ghailan A. Al Shafeey. (2019). A REVIEW OF SKIN DETECTOR BASED DEEP LEARNING TECHNIQUES: COHERENT TAXONOMY, OPEN CHALLENGES, MOTIVATIONS, RECOMMENDATIONS AND STATISTICAL ANALYSIS, FUTURE DIRECTION. *IEEE Access*. doi:10.1109/ACCESS.2019.2924989

Nada B. Ibrahim, M. Selim, & H. Zayed. (2012). A dynamic skin detector based on face skin tone color. *2012 8th International Conference on Informatics and Systems (INFOS)*.

Obeng, S. L., Danso, D. K., Omari, J. A., & Kuwornu-Adjaottor, J. (2018). Colour in Fashion: Effects on Personality. *European Journal of Education Studies*, 353-376. doi:10.5281/zenodo.1249188

P. Kakumanu, S. Makrogiannis, & N. Bourbakis,. (2007). A survey of skin-color modeling and detection methods. *Pattern Recognition Vol 40, Issue 3*, 1106-1122. doi:https://doi.org/10.1016/j.patcog.2006.06.010.

Pengzhi, L., Yan, P., & Jianqiang, L. (2023). A comprehensive survey on design and application of autoencoder in deep learning. *Applied Soft Computing*. doi:https://doi.org/10.1016/j.asoc.2023.110176

Perrett, D. I., & Sprengelmeyer, R. (2021). Clothing Aesthetics: Consistent Colour Choices to Match Fair and Tanned Skin Tones. *I-Perception*, 12(6).

Prof.Shivganga, G., Jayesh, P., Prajwal, T., Harshal, K., & Shusovan, M. (2020). Recommendation System using KNN and Cosine Similarity. *International Journal of Creative Research Thoughts (IJCRT), Volume 8*, 1255-1260.

Reeta, K., Anisha , G., Tejashri , W., & Siddesh , S. (2021). A Complexion based Outfit color recommender using Neural Networks. *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 1-7. doi:10.1109/ICAECT49130.2021.9392418

Ronny, V., & Na, L. (17 May, 2021). *Next-Generation Pose Detection with MoveNet and TensorFlow.js*. Retrieved from Tensorflow Blog: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html

Samantha Jackson. (26 August, 2019). *Building a Women’s Fashion Recommender*. Retrieved from Medium: https://medium.com/@sjacks/building-a-womens-fashion-recommender-2683856b97e3

Samit , C., Md. Saiful, H., Naimur , J. R., & Manik , B. C. (2021). Fashion Recommendation Systems, Models and Methods: A Review. *Informatics*.

Seema , W., Shruti, P., Pratik , S., Kriti , S., Mukund, K., C.V. Sri , V., & Ketan , K. (n.d.). Advanced Fashion Recommendation System for Different Body Types using Deep Learning Models. *Research Square*. doi:https://doi.org/10.21203/rs.3.rs-1856954/v1

Shaghayegh , S. (2021). *Image-based fashion recommender systems.* Luleå University of Technology, Department of Computer Science, Electrical and Space Engineering.

Shradha Dubey, & Manish Dixit. (2021). A comprehensive survey on human pose estimation approaches. *Multimedia Systems*. doi:https://doi.org/10.1007/s00530-022-00980-0

Shuangshuang, C., & Wei, G. (2023). Auto-Encoders in Deep Learning—A Review with New Perspectives. *Mathematics 2023, 11(8)*. doi:https://doi.org/10.3390/math11081777

Soham , N., Krish , P., Harsh , W., & Dr. Suvarna , P. (2022). Outfit Recommendation – Using Image Processing. *JOURNAL OF ALGEBRAIC STATISTICS*, 1699-1706.

Srinidhi , H., G.M. , S., & K.G. , S. (2022). Deep visual ensemble similarity (DVESM) approach for visually aware recommendation and search in smart community. *Journal of King Saud University - Computer and Information Sciences*, 2562-2573. doi:https://doi.org/10.1016/j.jksuci.2020.03.009.

Sudhir , K., & Mithun , G. D. (2019). c+GAN: Complementary Fashion Item Recommendation. *KDD ’19, Workshop on AI for fashion, Anchorage, Alaska-USA*. doi:https://doi.org/10.1145/nnnnnnn.nnnnnnn

T. J. McBride, N. Vandayar, & K. J. Nixon. (2019). A Comparison of Skin Detection Algorithms for Hand Gesture Recognition. *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*, 211-216. doi:10.1109/RoboMech.2019.8704839

W. Luo, & J. Xue. (2023). Human Pose Estimation Based on Improved HRNet Model. *2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI)*, 153-157. doi:10.1109/CCAI57533.2023.10201272.

Wang-Cheng , K., Chen Fang, Zhaowen , W., & Julian , M. (2017). Visually-Aware Fashion Recommendation and Design with Generative Image Models. *2017 IEEE International Conference on Data Mining (ICDM)*, 207-216. doi:10.1109/ICDM.2017.30

Wazhakar, S., Patil, S., Pratik S. Gupta, Singh, K., Khandelwel, M., Vaishnavi, C., & Kotecha, K. (2022). Advanced Fashion Recommendation System for Different Body Types using Deep Learning Models. *Research Square*. doi:https://doi.org/10.21203/rs.3.rs-1856954/v1

Xiangnan, H., Lizi, L., Hanwang, Z., Liqiang, N., Xia, H., & Tat-Sheng, C. (n.d.). Neural Collaborative Filtering. *WWW '17: Proceedings of the 26th International Conference on World Wide Web*, 173-182. doi:http://dx.doi.org/10.1145/3038912.3052569

YASHAR , D., FATEMEH , N., ARNAU , R., JULIAN , M., GIOVANNI , P., ALEJANDRO , B., & TOMMASO , D. (2022). A Review of Modern Fashion Recommender Systems. *ACM Computing Surveys*, 38. doi:https://doi.org/10.1145/nnnnnnn.nnnnnnn

Ying, H., & Tao, H. (2017). Outfit Recommendation System Based on Deep Learning. *Semantic Scholar*. doi:10.2991/ICCIA-17.2017.26

YuanZhe, P. (2022). A SURVEY ON MODERN RECOMMENDATION SYSTEM BASED. *arxiv: Information Retrieval*. doi:https://doi.org/10.48550/arXiv.2206.02631

Zatolokina, L. (13 August, 2023). *Using Human Pose Estimation in Fitness & Rehab Therapy Apps*. Retrieved from mobidev: https://mobidev.biz/blog/human-pose-estimation-technology-guide

Zhao, L., Jianke, Z., Jiajun, B., & Chun, C. (2015). A survey of human pose estimation: The body parts parsing based methods. *Journal of Visual Communication and Image Representation*, 10-19. doi:https://doi.org/10.1016/j.jvcir.2015.06.013

Ziwei, L., Ping, L., Qiu, S., XiaoGang, W., & Xiaoou, T. (2016). DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

# Appendix A: Literature Review Matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Authors (Year)** | **Focus** | **Methodologies** | **Key Findings** | **Limitations** | **Dataset Used** | **Considerations for Future Research** |
| Jaechoon, et al. (2020) | Fashion product retrieval model | Sketch-Product and Vector-Based User Preferred Fashion Recommendation Models | Developed models for product retrieval and personalized recommendations based on sketched images. | Not specified | Professionals filtered and collected with implicit user fashion profiling method. | Real-world implementation and consumer purchasing pattern analysis. |
| Soham, et al. (2022) | WebApp for matching outfits | CNNs, Agile methodology | Efficient WebApp for recommending outfits with shopping links | No metrics mentioned | Not specified | Review of the datasets used in fashion recommendation-based research publication |
| Aneesh K, et al. (2022) | Three-stage fashion recommendation system | CNNs for color and clothing type identification; Recommendation algorithm based on type, color, and occasion. | System for outfit combination suggestions | Details of recommendation algorithm is not specified. | Not specified | Incorporating user wardrobe items and broader style matrices |
| Guan, C., et al. (2018) | Knowledge-based apparel recommendation system | Multi-task CNN, SVM, and LKF classifiers | High prediction accuracy in recognizing clothing features and styles | Tailored for Menswear only | ATTRIBUTE and MEANING datasets from professionals | Expansion to more apparel categories and style meanings |
| M Sridevi, et al. (2020) | CNN-based fashion recommendation system | Transfer learning from ResNet50; Annoy library for similarity matching | High accuracy in fashion product recommendations | Limited to Rent the Runway's inventory | DeepFashion dataset | Diverse image sets and real-world captures for robustness |
| Sudhir & Mithun (2019) | Enhanced c+GAN for fashion recommendation | c+GAN model with MSE and perceptual loss; Clustering techniques | Improved 'Goes Well With' feature for garment matching | Not specified | 100k images scarped from Bing images portal | 3D modeling and user feedback for personalization |
| Wang-Cheng, et al. (2017) | Visually aware fashion recommendation system | DVBPR with CNNs; GANs for image generation | Combined visual feature extraction with personalized ranking and image generation | Not specified | 1 million web crawled images from amazon | Image generation enhancement for personalized design that includes fine-grained styles. |
| Atharv, et al. (2020) | User attribute-based fashion recommendation system | CNNs & RNNs, to extract features. | Personalized recommendations based on various user attributes (colour, age, gender, skin tone) | Require manual input of user attributes such as skin tone. | Not specified | Refining recommendation models for precision |
| Seema, et al. | Fashion recommendation tailored to body shape | Xception model; Body shape categorization | Accurate garment recommendations for different female body types | Women's apparel only | In-house curated dataset of women's apparel | Wider range of apparel and further accuracy improvements |
| Digant, et al. (2019) | Personalized fashion recommendation system | k-means for skin tone extraction; Decision tree algorithm for outfit selection | Personalized outfit recommendations based on skin tone and other parameters | Lack of evaluation metrics clarity | Manual dataset of 1030 outfits | Dataset expansion and algorithm refinement for precision |
| Reeta, et al. (2021) | “Pocket Fashionista” recommendation system | OpenCV for skin tone classification; CNNs and VGG16 for recommendations | Multiple modules for personalized outfit selection | Accuracy varies with photo quality and lighting | DeepFashion dataset | Classification accuracy and recommendation diversity improvements |