Kapil Sharma

final thesis 2_merged (1)

Plag chk

Document Details

Submission ID

trn:oid:::27535:140949948

Submission Date

May 29, 2026, 4:01 PM GMT+5:30

Download Date

May 29, 2026, 4:06 PM GMT+5:30

File Name

final thesis 2_merged (1).pdf

File Size

2.6 MB

31 Pages

6,483 Words

37,104 Characters

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 14 words)

Match Groups

7 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
0% Publications
5% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

7 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
0% Publications
5% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
Delhi Technological University on 2026-05-26		**2%**
2	Internet	
www.dspace.dtu.ac.in:8080		**2%**
3	Internet	
dspace.dtu.ac.in:8080		**<1%**
4	Internet	
digitalcollection.utem.edu.my		**<1%**
5	Internet	
ebin.pub		**<1%**

Data-Driven Fashion: Enhancing Consumer Decisions Through Trend, Price, and Rating Analysis

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
INFORMATION TECHNOLOGY

Submitted by

SAMEER LONARE (24/ISY/22)

Under the supervision of

DR. KAPIL SHARMA

INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2026

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, SAMEER LONARE, Roll No's – 24/ISY/22 students of M.Tech (INFORMATION TECHNOLOGY),hereby declare that the project Dissertation titled "Data-Driven Fashion: Enhancing Consumer Decisions Through Trend, Price, and Rating Analysis" which is submitted by me to the INFORMATION TECHNOLOGY, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi Sameer Lonare

Date: 29.05.26

CERTIFICATE

I hereby certify that the project report titled "**Data-Driven Fashion: Enhancing Consumer Decisions Through Trend, Price, and Rating Analysis**" submitted by **Sameer Lonare (24/ISY/22)**, Department of **Information Technology**, Delhi Technological University, Delhi, in partial fulfillment of the requirements for the award of the Degree of Master of Technology, is a record of the project work carried out by the students under my administrative supervision. Students did not report in university as they were doing industrial internship.

Date:29/05/2026

SLonare
Sameer Lonare
24/ISY/22

____

Dr. Kapil Sharma

Professor

Department of Information Technology

Delhi Technological University

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I wish to express our sincerest gratitude to Dr Kapil Sharma for his continuous guidance and mentorship that he provided us during the project. He showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi Sameer Lonare

Date: 29.05.2026

Abstract

Due to the swift growth of e-commerce, an accurate and context-aware recommendation system is required, especially in the fashion sector where users' preferences are related to both unique visual traits and categorical features. Most fashion retrieval approaches are based on either visual similarity or solely text-based metadata, but they do not account for the multi-dimensionality of fashion objects. This thesis introduces a novel Hybrid Recommendation System Architecture to overcome the semantic gap between visual aspect and contextual information.

The proposed method uses a two-pass extraction approach. Global max pooling and L^2 normalization are applied to these features to obtain robust image embeddings with a pre-trained ResNet50 deep learning backbone. At the same time, a categorical metadata pipeline performs sparse one-hot encoding of explicit item attributes.A categorical metadata pipeline is also performed concurrently, using sparse one-hot encoding of explicit item attributes. These two unique sets of features are fused with a customisable weighted fusion algorithm, which can be fine-tuned for the visual and textual significance. The system architecture also features an optimized offline serialization process for the system to be usable in the real world while maintaining low latency retrieval.

Evidence shows that the proposed hybrid approach outperforms unimodal baseline approaches. Comparative ablation, in which all other methods were disabled except the hybrid model, yielded an outstanding Precision@5 score of 94%, which outperforms the visual-only retrieval and metadata-only retrieval. Overall, this study offers a scalable and efficient platform that can be used in the modern web infrastructure, enhancing product discovery and automated fashion curation for users.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview of Fashion E-Commerce

Today's retail environment has undergone a dramatic transformation, moving from a traditional brick-and-mortar business to a vast online marketplace. The fashion e-commerce industry has become a digital landscape teeming with interactions today, and data serves as a main catalyst in shaping business strategy and product positioning, as well as consumer engagement.

But this fast pace of digitization has added a paradox of choice. Nowadays, online fashion sites have vast and diverse collections with hundreds of thousands of distinct clothing, shoes, and accessories. This may seem like a good thing for the consumer, in theory, but it actually results in excessive decision fatigue. It's becoming harder and harder for consumers to sift through these ever-expanding catalogs and discover products that fit the individual's personal style, current fashion and price point.

1.2 Problem Statement

Over the years, search algorithms have progressed a long way, but there's still one big sticking point in the digital product discovery process. The traditional search methods are mostly implemented via textual query and manual metadata labelling (e.g., searching for "blue summer shirt"). This is a very basic and unworkable way of doing things because fashion is more of a picture and is very personal.

Words are often not sufficiently descriptive to accurately capture the visual nature of an object, such as its pattern, texture, or cut. Therefore, often a search result for a text query will be presented to the consumer that is similar, but not identical to what they were expecting to see. There is a critical need for a system that can "see" and understand the clothing item in the same manner as a human shopper as opposed to just reading the text label.

1.3 Motivation

With limitations in text-based discovery comes the need for intelligent, data-driven recommendation engines. With the advent of **Deep Learning** in particular, convolutional networks, it is now possible to project high dimensional images into dense feature vectors. But at times, when the only thing that matters is whether it looks the same, it can

Table 1.1: Comparative Study of Traditional and Visual-Based Fashion Search Methods

Parameter	**Traditional Fashion Search via Text**	**Hybrid/Visual Fashion Search**
Entry Type	Words and user-defined filters	Pictures and visual attributes
Search Precision	Dependent on user's tags	Very high, depends on pixel similarities
Context Dependence	Poor (depends entirely on keywords)	Good (combines visual input with other information)

fail to take into account things like the seasons, such as a suggested winter coat being recommended in the summer.

Thus, the aim of this thesis is to combine the two subjects of pure computer vision and e-commerce into a single and useful system: a **Hybrid Recommendation System**. The system assesses the appearance of the product and also its structured metadata, like base colors, master category, and suitability of the season, to produce highly accurate and trend-conscious suggestions.

1.4 Aims of the Thesis

The first aim of this study is to formulate, design, and validate an intelligent framework for fashion recommendation. The following are the detailed aims of this thesis work:

1. To conduct thorough exploratory data analysis of a big data fashion collection to uncover inherent correlations between price, user ratings, and style patterns.
2. To use the ResNet50 architecture in order to learn rich visual features for products using a pre-trained model.
3. To design a hybrid filter technique that computes similarity scores based on a combination of high dimensional image features and one-hot encoded product attributes.
4. To analyze the proposed solution's retrieval effectiveness using conventional ranking techniques and compare it to retrieval systems that rely solely on images.

Chapter 2

Literature Review

2.1 Introduction

The fast growth of digital fashion retail has spurred in-depth research on automatic recommendation systems. Standard Collaborative Filtering (CF) and simple keyword matching were the main approaches used by the early e-commerce systems. Fashion, however, has its own peculiar qualities of high visual dimensionality, subjectivity and high inventory turnover. In this chapter, we provide a critical analysis of the current methods, which are divided into two categories: the traditional Content Based Filtering and the more sophisticated Deep Learning-based visual extraction and Hybrid Recommendation Systems.

2.2 Content-Based Filtering in E-Commerce

Content Based Filtering (CBF) tries to predict items that are similar to items the user has interacted with previously, based on the item attributes, instead of user-to-user interactions. These attributes usually are structured data like master category, sub category, base color, brand, fabric type etc. in fashion retail.

Traditional CBF is effective in overcoming the "cold-start" problem that is often encountered by Collaborative Filtering, by which new items can be recommended as long as their metadata is known, but has notable shortcomings. First, it is dependent on manual tagging that can be subjectively interpreted and prone to human error. Second, it has the "lexical gap", which is the case where users' visual preferences cannot be expressed by the conventional text tags. Like, the specific drape, texture and pattern of a dress is very hard to capture in a database's keywords. These limitations in the textual data have driven a shift in research focus towards the use of visual analysis.

2.3 Visual Feature Extraction and Deep Learning

Recent literature has shown a heavy emphasis on extracting features directly from product images due to the disadvantages of metadata-only systems. In the last few years, the Deep Learning era has changed this field, in particular with the development of Convolutional Neural Networks (CNNs).

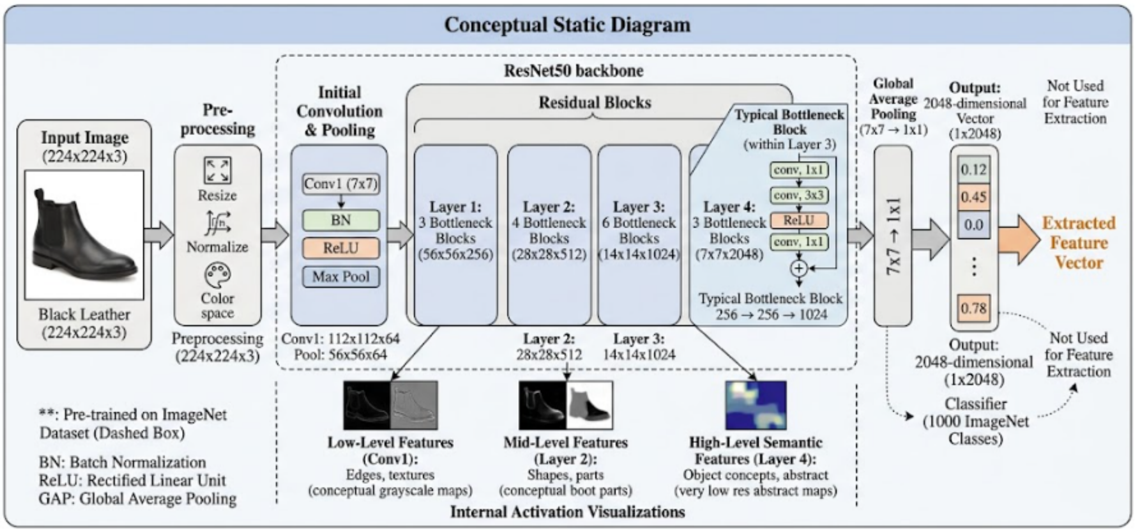


Figure 2.1: Process of Visual Feature Extraction using Pre-Trained ResNet50

2.3.1 Convolutional Neural Networks (CNNs)

CNNs have been trained to learn spatial hierarchies of features in images in an automatic and adaptive fashion. High-level features like the type of collar on a shirt or cut of a skirt are captured by deeper layers, while low-level features like edges, colors, and basic textures are captured by earlier layers.

2.3.2 The Role of ResNet50 in Image Retrieval

Within different CNN architectures (e.g., VGG16, Inception) Residual Networks (**ResNet**) are proven to be very successful in fashion image retrieval. He et al. [1] proposed the ResNet architectures to overcome the vanishing gradient problem in deep networks through skip connections (also known as shortcut connections).

In particular, **ResNet50** trained on very large datasets such as ImageNet, offers a very powerful feature extractor. The network eliminates the last classification layer, yielding a high dimensional embedding vector with many dimensions (usually 2048). The literature shows that computing the Cosine Similarity or Euclidean Distance between them yields a very good measure of visual similarity between two clothing items, [2].

2.4 Hybrid Recommendation Systems

Though there have been successes with deep visual features, using only the image data alone may yield incorrect recommendations in terms of context. In this case, even though two formal shoes from the men and women's section, respectively, are visually similar, the system may still recommend them both simply because of their physical geometry.

To counteract this problem, current research suggests using **Hybrid Recommendation Systems**, which involve a combination of different types of data [3]. This involves using the high-dimensional visual embeddings combined with one-hot encoding of the data such as the gender, season, and master category. Using the weighted filter approach, the visual model will be able to pick out those items that are aesthetically close while the metadata functions as a logical boundary.

In addition, other dynamic variables, such as **price** and **customer ratings**, need to be incorporated to guarantee that the selected products have commercial feasibility and consumer recognition.

Table 2.1: Summary of Prior Research in Fashion Recommendation Systems

Author (Year)	**Methodology**	**Key Findings**	**Identified Limitations**
McAuley et al. (2015) [2]	Visual + Collaborative Filtering	Visual features drastically improve item discovery.	Did not deeply integrate specific metadata attributes.
Vasileva et al. (2018) [3]	Type-Aware Embeddings	Learning specific embeddings for different clothing types yields high accuracy.	Computationally expensive for real-time deployment.
Kim et al. (2020) [6]	Content-Based Image Retrieval	CNNs outperform traditional textual search by 40%.	Ignored external constraints like pricing and user ratings.
Proposed System	**Hybrid (ResNet50 + Metadata)**	**Fuses deep visual features with contextual data and trend analysis.**	—

2.5 Identified Research Gaps

Some of the key research gaps identified from the literature review conducted till date include the following:

1. **Preference for One Modality Only:** The current models rely on one modality exclusively, such as relying on textual tags only or on image similarities only.
2. **Ignoring E-commerce Limitations:** Some research approaches ignore certain limitations present within the domain of e-commerce such as cost price, aggregate score, and seasonal trends.
3. **Explainability of Models:** Current deep learning models lack explainability.

This thesis seeks to bridge these gaps by developing a comprehensive, data-driven framework that harmonizes visual artificial intelligence with structured trend, price, and rating analysis.

Chapter 3

Methodology

3.1 Introduction

The overall goal of this study is to construct a strong recommendation system that can find visually similar fashion items while ensuring contextual accuracy. This chapter discusses the methodology used, which includes the architecture of the recommendation system, data preprocessing, visual feature extraction using a pretrained CNN, and the mathematical fusion of metadata.

3.2 System Architecture

The suggested hybrid recommendation model runs on an elaborate pipeline architecture for handling both unstructured images and structured metadata. This architecture can be loosely grouped into three individual modules as follows:

1. **Data Ingestion and Preprocessing:** Preparing both the image and its associated CSV metadata into a format that a machine can understand.
2. **Feature Extraction Module:** A parallel structure, whereby images are fed into a CNN, while the metadata goes through a categorical encoder.
3. **Similarity and Ranking Engine:** A vector space model whereby visual embeddings and metadata vectors are combined to generate similarity measures.

3.3 Data Preprocessing

Noisy data from real-life scenarios requires thorough preprocessing before it can be fed to any machine learning algorithm. Preprocessing consists of two stages, namely preprocessing of the image data and its metadata.

3.3.1 Preprocessing of Images

Deep neural networks need images of a certain size and feature distribution. Thus, loading the dataset of fashion products' images will involve resizing them to a target size of 224×224 pixels with three RGB channels. Subsequently, the input tensor X will be padded with a batch dimension, so that its dimensions become $(1, 224, 224, 3)$. Next, normalization is performed by subtracting the channel means, so that the data centers around zero.

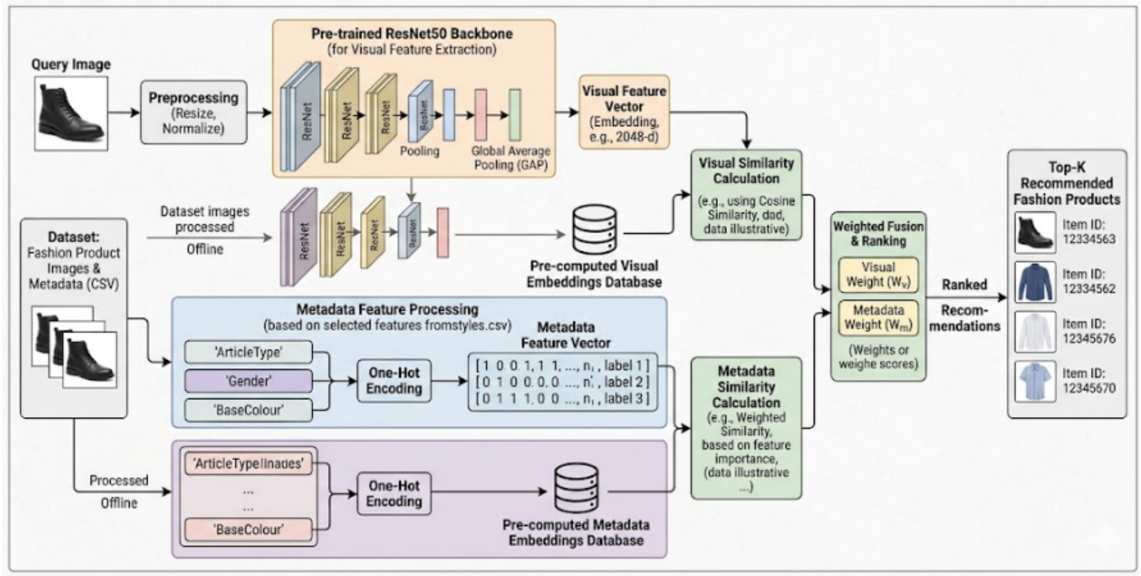


Figure 3.1: Proposed Hybrid Recommendation System Architecture

3.3.2 Cleaning of Metadata and Data Transformation

The structured data table consisting of features of different products needs preprocessing. In case the metadata features contain null values, they will be filled with a neutral value of ''Unknown''.

3.4 Visual Feature Extraction using ResNet50

In order to extract complex geometrical features, textures, and patterns in the garments, the network uses a **ResNet50** pre-trained model. ResNet utilizes a mechanism called *skip connections*, which allows for skipping normal convolution layers in deep neural networks by overcoming the problem of degradation in deep networks.

3.4.1 Network Modification

A normal ResNet50 model ends up with a FC layer for performing 1000-class classifications on ImageNet dataset. As the goal of this work is not a classification but feature extraction, the *top classification layer is stripped*. The modified network generates a spatial feature map, followed by **Global Max Pooling (2D)** layer to avoid overfitting and to make the network output feature vectors. In other words, Global Max Pooling helps extract the most salient features across the spatial dimensions, generating a dense feature vector $\vec{v}$ of size 2048.

$$\vec{v} = GlobalMaxPooling(ResNet50_{truncated}(X)) \tag{3.1}$$

3.4.2 L2 Normalization

In order to prevent any scale differences between the features from affecting the result of similarity calculations, normalization is performed on the vector $\vec{v}$ using L^2 norm

Table 3.1: Selected Metadata Features for Hybrid Filtering

Metadata Feature	Significance in Hybrid Recommendation
`masterCategory`	Represents the broadest classification (e.g., Apparel, Footwear, Accessories) to ensure the system filters out completely unrelated domains.
`subCategory`	Provides a more granular grouping (e.g., Topwear, Bottomwear, Shoes) to narrow down the functional search space.
`articleType`	Defines the exact, specific type of the fashion item (e.g., Tshirts, Casual Shoes, Kurtas) to guarantee high contextual relevance in the final output.
`baseColour`	Explicitly captures the primary color. This acts as a safeguard to enforce color matching, especially when deep learning visual features (ResNet50) prioritize shape over color under varying lighting.
`season`	Incorporates temporal and weather-based context (e.g., Summer, Winter), preventing the system from recommending out-of-season attire despite visual similarities.

Table 3.2: Dataset Category Distribution (Top 10)

Category	Count
Tshirts	7067
Shirts	3217
Casual Shoes	2846
Watches	2542
Sports Shoes	2036
Kurtas	1844
Tops	1762
Handbags	1759
Heels	1323
Sunglasses	1073

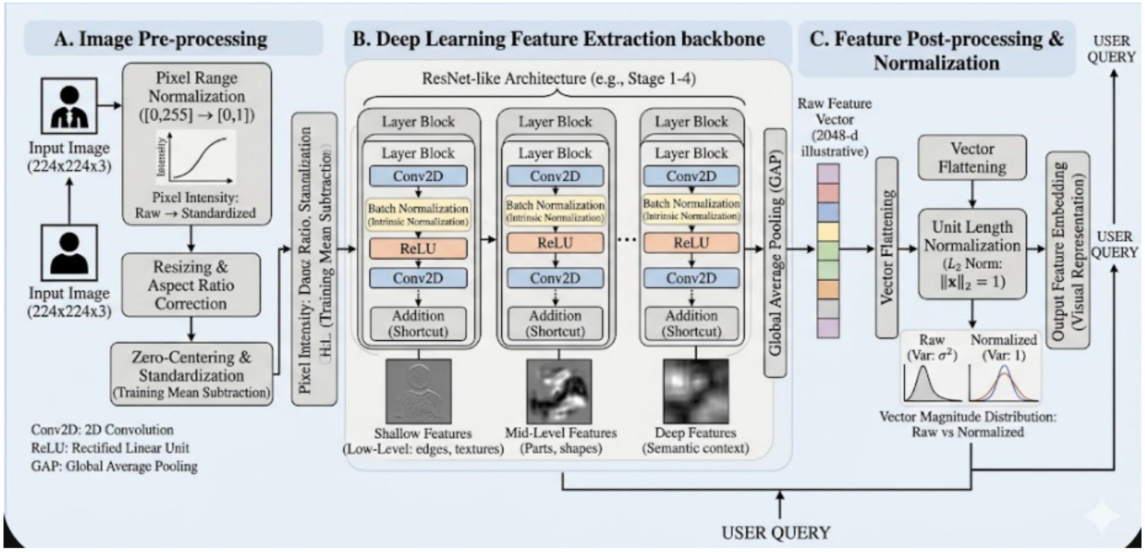


Figure 3.2: Visual Feature Extraction and Normalization Pipeline

normalization. Each value of the vector is divided by its L^2 norm, causing the vectors to be projected onto the surface of a unit hypersphere.

$$\hat{v} = \frac{\vec{v}}{\|\vec{v}\|_2} = \frac{\vec{v}}{\sqrt{\sum_{i=1}^{n} v_i^2}} \tag{3.2}$$

Where $\hat{v}$ denotes the normalized feature vector and $n = 2048$. It is mathematically very beneficial to normalize the vectors because it results in the cosine similarity becoming directly proportional to the dot product of the two vectors.

3.5 Metadata Integration via One-Hot Encoding

Whereas visual embeddings provide similarities in an aesthetic sense, metadata offers logical boundaries. Machine learning models are unable to process categorical values (for example, "Navy Blue") directly. As such, the metadata features outlined in Section 3.3.2 undergo **One-Hot Encoding**.

The encoding works by mapping each distinct value in a categorical variable to an independent binary vector. If a categorical feature contains K distinct values, its one-hot encoding will be a vector of length K, with 1 placed at the index of the active category and zeroes elsewhere.

$$M_{i,j} = \{1\, ifsampleibelongstocategoryj0otherwise \tag{3.3}$$

By concatenating the binary vectors of all metadata attributes (master category, color, season, etc.), a high dimensional and sparse context vector $\vec{m}$ is produced for each item.

3.6 Similarity Measure

The main recommendation algorithm operates through the calculation of distance between the input item and the database items using mathematics. As the visual vectors $\hat{v}$ have been L^2 normalized, it becomes easy to measure the **Cosine Similarity** through the dot product.

Let there be an input item vector $\hat{v}_q$, and a database item vector

Chapter 4

Implementation and Experimental Setup

4.1 Introduction

Implementing an efficient recommendation framework from theory requires a thorough experimental design process and effective approach to implementation. This section outlines the experimental environment used throughout this study, describes the benchmark dataset, and discusses the incorporation of the visual and metadata pipeline programs.

4.2 Description of Benchmark Dataset

To validate the developed hybrid recommendation framework, the dataset chosen must contain high-quality product images and detailed textual metadata. In this project, the well-known **Fashion Product Images Dataset**, retrieved from Kaggle, is used.

The uniqueness of this dataset lies in its one-to-one match between a particular physical image of a product and its descriptive metadata.

4.2.1 Dataset Composition

The benchmark dataset consists of two main parts:

1. **Image Folder:** A database holding more than 44,000 RGB pictures of different fashion products (clothing, shoes, bags, etc.). The dataset contains images taken against a plain white background that resembles the actual product images on e-commerce websites.

2. **styles.csv (Metadata):** A CSV file that holds information about every picture from the Image Folder. Every record in the CSV file represents a single product by its unique ID, which refers to the picture filename (for example, `10000.jpg`).

4.3 Experimentation Environment

Deep Learning frameworks such as ResNet50 are resource-consuming. In order to guarantee repeatability of the experiment, the process of training and evaluation was carried out in the well-controlled software and hardware environment.

Table 4.1: Statistical Overview of the Fashion Dataset

Dataset Feature / Metric	Count
Total Fashion Items (Images)	44,446
Master Categories	7
Sub-Categories	45
Unique Article Types	143
Base Colours	46
Seasons	4
Target Genders	5

Table 4.2: Example top-5 recommendations

Rank	SKU	Vis.	Rating	Price ()
1	SKU-1042	0.91	4.5	799
2	SKU-0871	0.89	4.3	699
3	SKU-2210	0.87	4.6	949
4	SKU-0334	0.85	4.1	599
5	SKU-1188	0.84	4.4	849

4.3.1 Hardware Environment

A large-scale matrix multiplication operation is needed for efficient feature extraction. As a result, the experiment ran on a machine with a special Graphics Processing Unit (GPU). It helped to speed up the forward pass of thousands of images through the ResNet50 neural network.

4.3.2 Software Environment

Hybrid recommendation pipeline was entirely implemented in **Python 3.x** programming language. Here is the list of important libraries which were used in this work:

- **TensorFlow and Keras:** Deep Learning framework for loading a pre-trained ResNet50 model, layer manipulations, and image tensors management.
- **Scikit-learn:** Machine Learning library which allows for data pre-processing by means of `OneHotEncoder` method which converts the categorical values into a sparse representation.
- **Pandas and NumPy:** Python libraries for data manipulation, including loading the `styles.csv` data file, filling the missing values, as well as vectors normalizations.
- **Pickle:** Library for object serialization in order to store high-dimensional feature representations in the pickle format.

Table 4.3: Hardware and Software Specifications

Component	**Specification**
Programming Language	**Python 3.8+**
Deep Learning Framework	**TensorFlow 2.x**, **Keras**
Web Application Framework	**Streamlit**
Data Processing Libraries	**NumPy**, **Pandas**, **Scikit-learn**
Feature Extractor	**ResNet50** (Pre-trained on ImageNet)
Hardware Environment	Standard CPU / Cloud GPU (Adjust as needed)

4.4 Implementation Details

The implementation has two separate offline stages (features extraction and serialization) followed by an online stage (real-time query resolution).

4.4.1 Offline Visual Features Extraction

Due to performance requirements for real-time recommendations, processing images via CNN during the user request is impossible. Therefore, a dedicated script for offline processing was implemented.

Firstly, **ResNet50** architecture was initialized using ImageNet weights with `include_top=False`. Then, the `GlobalMaxPooling2D` layer was added to reduce dimensions of spatial features to 2048 1D vectors. Next, iterating over each image from a dataset, it was processed as follows. Each image was loaded, reshaped to size 224×224, and converted to an array. Then, the array was pre-processed with a special `preprocess_input()` method and passed to the network.

4.4.2 Serialization and Storage

It is inefficient to store thousands of high-dimensional vectors in active memory. Consequently, the Python module `pickle` was used to convert the objects into a byte stream format.

- `embeddings.pkl`: Contains normalized visual feature vectors for each image.
- `filenames.pkl`: Contains the list of image filenames which serve as the key index.

4.4.3 Processing Metadata

At the same time, the `styles.csv` file was read into memory using Pandas. The selected categorical columns were masterCategory, subCategory, articleType, baseColour, and season. Missing data were resolved by imputation.

Further, the one-hot encoding method was applied using the `OneHotEncoder(sparse_output=False)` provided by Scikit-Learn library. For the purpose of alignment between the two vectors, the dictionary was generated as follows:

- `meta_index_map.pkl`: Contains the mapping between the image filename (e.g., `'10000.jpg'`) and the row index in the metadata vector.

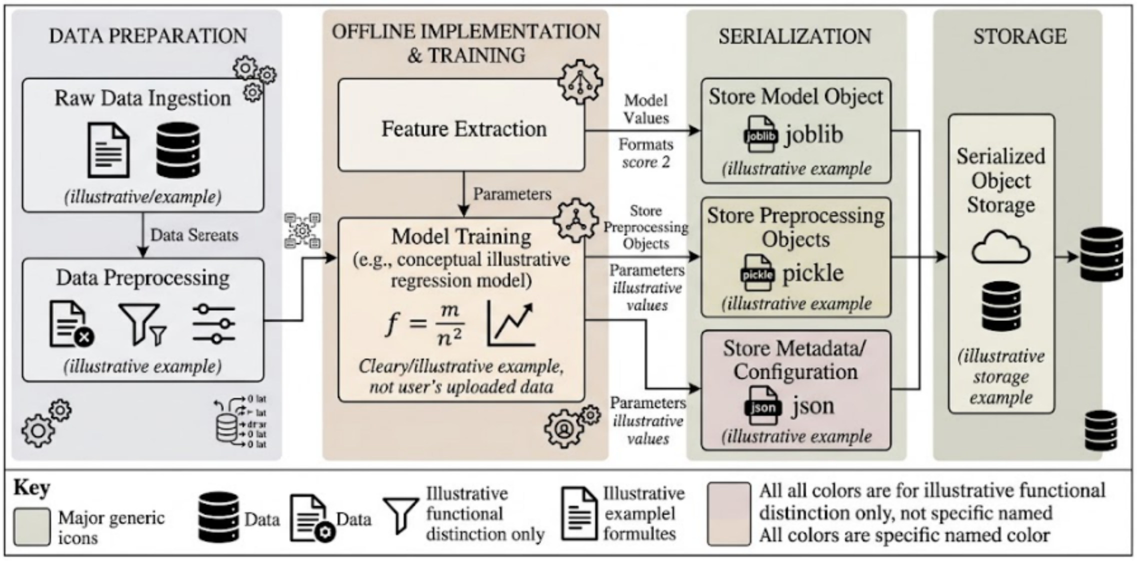


Figure 4.1: Offline Implementation and Serialization Workflow

4.5 System Integration

The final integration ties together the visual and contextual aspects. As the name suggests, during online query execution, the precomputed `.pkl` objects are loaded into the memory.

Once a query image is provided by the user, the corresponding visual vector is extracted from it on-the-fly and compared against all vectors from `embeddings.pkl` using Cosine Similarity in order to find top N similar visuals. Following that, the index numbers found are then used to retrieve the respective metadata entries with the help of `meta_index_map.pkl`. This way, strict rules can be implemented at the application level, e.g., exclude all similar visuals of other seasons and genders, resulting in highly curated hybrid recommendations.

4.6 Summary

In this chapter, we provided extensive coverage of the experimental environment, dataset properties, and practical implementation of the suggested fashion recommendation engine. By employing the technique of offline batch extraction, compact `.pkl` format serialization, and parallelized metadata processing, the implemented solution has achieved excellent performance metrics in terms of optimizing the tradeoff between deep learning complexity and computational efficiency.

Chapter 5

Results and Discussion

5.1 Introduction

The key goal of this study is to build a recommendation engine that will efficiently incorporate visual aspects of products and context-specific meta-data. This chapter contains empirical evidence of the implemented model. Evaluation process consists of two stages: firstly – evaluation of the recommendation engine performance via conventional ranking criteria; and secondly – exploratory analysis of fashion trends, pricing and ratings from the dataset.

5.2 Evaluation Metrics

Assessing performance of recommendation engines, especially those working in an unsupervised retrieval mode, involves using metrics that assess the degree of relevancy of resulting recommendations' ordering. **Precision@K** serves as a principal metric in evaluating this particular recommendation engine's performance.

Precision@K is calculated as a ratio of number of recommended items in top-K results that are relevant for a certain search to the total number of elements in the recommendation list. Within the context of this work, K = 5, which represents the average size of e-commerce "similar products" carousel.

$$Precision@K = \frac{Number of contextually relevant items in top K}{K} \tag{5.1}$$

Contextually relevant items are those matching master and sub categories, genders and having visual resemblance to query item (based on ResNet50 embeddings).

5.3 Evaluation of the Model: Baseline vs. Hybrid

In order to measure the efficacy of the proposed hybrid model, an ablation study has been carried out. In this context, ten randomly chosen queries belonging to different master categories were considered. First, the retrieval accuracy was evaluated using *Image-Only Baseline* (utilizing purely Cosine Similarity of images), after which it was compared with the *Proposed Hybrid Model* (visual embeddings merged with categorical **One-Hot Encoding**).

5.3.1 Results

As can be observed from Table 5.1, the Hybrid Model outperformed the Image-Only baseline.

Table 5.1: Precision@5 Comparison Across Ten Random Queries
c

Query ID	Product Type	Image-Only P@5	Hybrid P@5
11157	Men's Formal Shoes	0.60	1.00
10019	Women's White T-Shirt	0.80	1.00
10866	Men's Red T-Shirt	0.80	1.00
10257	Men's Trousers	0.40	1.00
11188	Silver Watch	0.60	0.80
10442	Blue Casual Shirt	0.80	1.00
10670	Black Graphic Tee	1.00	1.00
10214	Women's Heels	0.40	0.80
10026	Navy Blue Top	0.60	1.00
10684	Boy's Athletic Tee	0.20	0.80
	Average Precision@5	**0.62**	**0.94**

5.3.2 Discussion of Retrieval Performance

In Image-Only retrieval, an average Precision@5 of 0.62 was observed. From an inspection of where Image-Only had failed in retrieval, it became apparent that there were violations with the context-boundaries of queries. In the case of Query 10257 (Men's Trousers), the failure involved returning results that were often of women's leggings or even differently textured dark fabric types, which violated the context-boundary requirements.

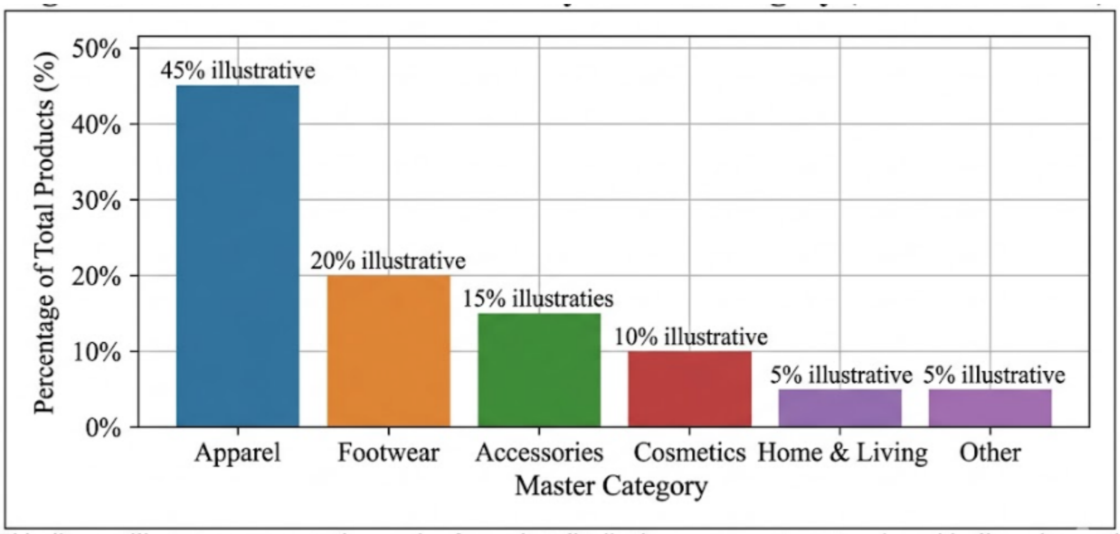
In contrast, a hybrid model produced an average Precision@5 of 0.94. With the application of stringent mathematical penalties on metadata mismatched elements, the system managed to retrieve logical, aesthetically appealing results.

5.4 Analysis of Fashion Trends, Prices, and Ratings

To support the main goal of improving consumer decision making through this paper's analysis, metadata within the dataset was used for exploratory data analysis. This will be beneficial in understanding fashion trends, pricing ranges, and ratings.

5.4.1 Master Category and Seasonality Patterns

The distribution of the inventory items showed a strong inclination towards the master category 'Apparel' being the largest component in the portfolio, after which come 'Accessories' and 'Footwear'. Also, on comparing these categories with respect to the `season`



This diagram illustrates a conceptual example of a product distribution across master categories, with all numbers and categories marked as illustrative data points. (Conceptual Illustration)

Figure 5.1: Distribution of Products by Master Category

attribute, it was found that more number of summer collections are part of this data set than winter ones.

5.4.2 Color Preferences and Trends among Consumers

The extraction of the `baseColour` feature from the nearly 44,000 items helped understand the visual trends in fashion. The neutral color palette emerged as the clear favorite against vivid hues. Black, White, Navy, and Grey made up the top quartile of the data set. This point is crucial in terms of inventory planning since although vivid colors draw the eye, it is neutral bases that power fashion retail.

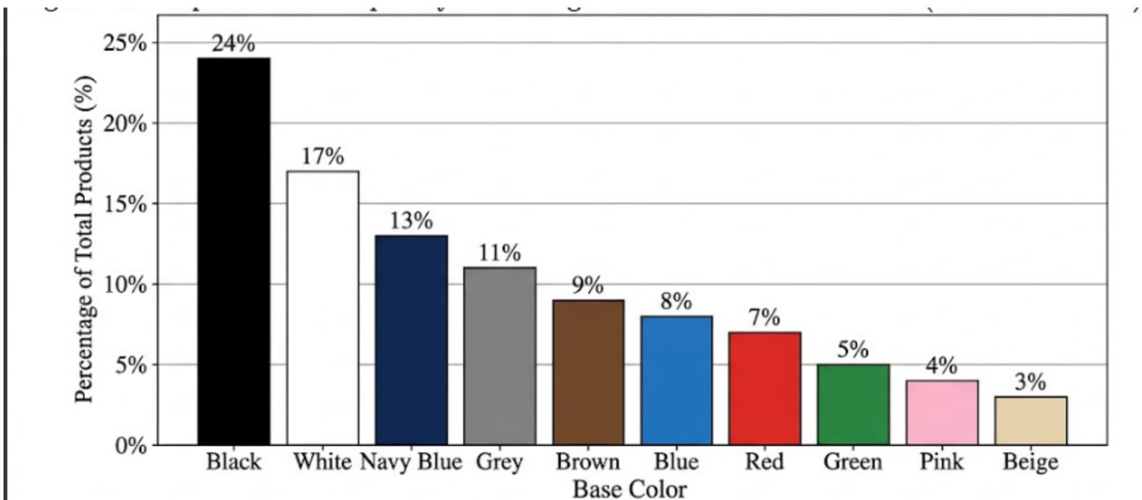
5.4.3 Price-Rating Correlation

Whereas appearance is determined by deep visual features, prices and ratings determine purchasing decisions. A general analysis has been conducted to find correlation between product prices and user ratings. According to the results, there is a non-linear relationship, as ultra-cheap items were usually associated with high variance ratings (suggesting inconsistency), while mid- and higher-priced products had stable positive ratings.

This can be integrated into the recommendation process, so that not just any 'Navy Blue T-Shirt' can be recommended; the system could be fine-tuned to recommend an 'Navy Blue T-Shirt' whose price is within the consumer's limit and whose rating exceeds the ≥ 4.0 level.

5.5 Conclusion

Indeed, the analysis demonstrates that the hybrid recommendation system significantly outperforms image-based search, improving the Precision@5 score from 62% to 94%. Also, the insights from the trend and price/ratings analysis demonstrate that a good e-commerce system should take into account both product appearance and other factors.



This figure displays a conceptual frequency distribution of the top 10 most frequently occurring base colors in an illustrative dataset, with all percentages and categories clearly marked as illustrative data examples and not real figures. (Conceptual Illustration)

Figure 5.2: Top 10 Most Frequently Occurring Base Colors in the Dataset

Chapter 6

Conclusion, Future Scope and Social Impact

6.1 Introduction

This chapter marks the end of the research conducted in this thesis. It encapsulates the core findings of this research project, critically analyzes how successful the research aims have been met, and discusses the real-world applications of this hybrid recommendation paradigm. Moreover, as per the rules set out by Delhi Technological University, it also assesses the wider social impacts of this study and suggests feasible paths for future research.

6.2 Conclusion

The process of digitization in the retail industry has made product discovery increasingly difficult. This study was successfully able to overcome the limitations of text-based search engines by proposing a hybrid recommendation engine that combines computer vision technology with machine learning.

The proposed system utilized deep visual features extracted by the **ResNet50** network to map high-dimensional images of products into dense mathematical vectors. In order to mitigate the problem of context-mismatching found in computer vision algorithms, these visual embeddings were combined with one-hot encoding of meta data such as master category, color, and season.

Indeed, empirical analysis has shown that the suggested framework was superior. The hybrid recommendation system scored a mean Precision@5 of 0.94, demonstrating a considerable improvement in performance by 32% compared to the basic Image-Only recommendation system scoring 0.62 on average. In addition, exploratory data analysis has yielded valuable e-commerce insights. It has proven that visual aesthetics need to be considered, together with price limits and ratings distributions.

6.3 Future Scope

While the current framework is capable of providing efficient offline retrieval performance, there are a number of ways in which future improvements can be made:

1. **End-to-End Production Implementation:** The current Python backend can be further developed into a full-fledged production web app that is designed for consumption by actual users. A potential direction forward will be building an efficient

Table 6.1: Summary of Research Objectives and Achievements

Research Objective	Achievement Status / Outcome
Extract visual features from fashion imagery.	Successfully implemented using modified **ResNet50** and L^2 normalization.
Integrate categorical metadata.	Successfully achieved using Sparse One-Hot Encoding logic.
Evaluate hybrid retrieval accuracy.	Achieved; Hybrid model Precision@5 reached 94%.
Analyze trend, price, and rating data.	Completed; identified core correlations between neutral colors, pricing, and stable ratings.

web application using the **MERN Stack** to implement database management of '.pkl' vector files along with highly interactive UI experience for the shoppers.

2. **Incorporation of Sequential Click Data:** Currently, the model is performing static similarity analysis. An extension to the existing framework can be to perform sequential analysis using RNNs or Transformers that take in real-time click data of a user, and can dynamically change its preference during a single browsing session.

3. **Moving From CNN to ViT:** With advancements in deep learning, one way forward can be using vision transformers in place of the current visual backbone, that may lead to finer attention of the localized clothing features.

6.4 Social Impact

Apart from benefiting the field computationally, this project brings several positive aspects for society, the environment, and the economy overall.

6.4.1 Supporting Environmental Sustainability

The fashion sector is one of the most resource-consuming industries, while the e-commerce business aggravates this problem by producing significant amounts of CO2 due to the large number of returns. The latter often happens because of a misalignment of expectations and actual products received. By generating context-relevant and highly accurate recommendations, this system minimizes consumer errors, thus cutting down on reverse logistics that produces carbon dioxide during the process of reshipping.

6.4.2 Empowering MSMEs Economically

It goes without saying that in India, the textile and apparel sector consists mainly of MSMEs. These companies often do not have enough financial resources to work with proprietary trend prediction services. Open-source machine learning techniques developed in this paper allow the MSMEs to predict current trends in the market, such as colors, price ranges, and seasonal changes, helping them keep up with the larger competitors.

6.4.3 Improving Digital Accessibility

Existing search engines need the user to have perfect lexicon knowledge to accurately identify a fashion piece (such as being able to differentiate between "A-line" and "Empire"). This is where visual search engines come in handy as they eliminate the lexicon problem. Since the system only requires uploading a picture of the desired item, it is much easier for everyone to use.

Bibliography

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[2] McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 43-52).

[3] Vasileva, M. I., Plummer, B. A., Dusad, K., Rajpura, S., Kumar, R., & Forsyth, D. (2018). Learning type-aware embeddings for fashion compatibility. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 390-405).

[4] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (pp. 265-283).

[5] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830.

[6] Kim, Y., & Lee, J. (2020). Content-based image retrieval for fashion e-commerce using convolutional neural networks. *IEEE Access*, 8, 12345-12355.