Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.46
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Full Length Research Article

A HYBRID DEEP NEURAL ARCHITECTURE FOR PERSONALIZED SKINCARE RECOMMENDATION

Philip O. Odion and *Khadijah Nasir Atiku

Department of Computer Science, Faculty of Military Science and Interdisciplinary Studies, Nigerian Defence Academy NDA, Kaduna

*Corresponding Author Email Address: khadijahnasir757@gmail.com

ABSTRACT
Personalized skincare recommendations require accounting for both user preferences and the biological suitability of products. This study investigates personalized skincare recommendations through a neural model that integrates user behavior and ingredient-level product information. A Multi-Head Attention Two-Tower neural architecture was developed that combined user latent embeddings with product embeddings derived from ingredient metadata. The model was trained using ranking-based objectives on the Amazon Beauty dataset with 5-core filtering, comprising 5,269 user–item interactions. Performance was evaluated using standard ranking metrics, including Hit Rate (HR@K), Recall@K, and nDCG@K. The model achieved a Hit Rate of 12.41% at top-10 recommendations, outperforming random selection by an order of magnitude. Stratified analysis by skin type revealed the highest effectiveness for sensitive skin (HR@10 = 21.9%) and the lowest for dry skin (HR@10 = 3.6%), which demonstrates the system's ability to encode biologically relevant compatibility patterns. Rank distribution analysis confirmed that recommendations were not dominated by popularity, with relevant items consistently prioritized across users. The results confirmed that incorporating attention mechanisms with a dual-tower architecture enables biologically informed and personalized recommendations. The findings support integrating explicit skin profiles and ingredient-level embeddings to enhance safety, relevance, and user trust in skincare recommendation systems.

Keywords: Personalized recommendation systems; Two-Tower neural network; Attention mechanisms; Skincare recommendation; Deep learning; Ingredient-aware recommendation.

INTRODUCTION

The global economy has undergone extensive transformation due to the rise of e-commerce (Hidayat et al., 2025). Over the past few decades, digital platforms have redefined how individuals discover, evaluate, and purchase goods (Chouhan & Barde, 2024). What began as a novel experiment in online retail has become one of the most powerful drivers of global trade. Platforms such as Amazon and Sephora offer thousands of alternatives that differ in formulation, ingredient composition, and targeted skin concerns (Khan & Singh, 2025). While this abundance expands consumer choice, it also increases decision complexity, especially for individuals with specific skin conditions such as acne, dryness, or sensitivity (Dinata & Baizal, 2024). Inappropriate product selection may result in irritation, allergic reactions, or worsening dermatological problems. Personalized recommendation systems are increasingly viewed as essential tools for assisting consumers in identifying suitable skincare products (Utami et al., 2025). Recommender systems have been widely adopted in e-commerce to predict user preferences based on historical interactions. Early approaches relied on collaborative filtering and matrix factorization techniques to model user–item relationships (Han, 2020). More recent work incorporates deep learning architectures to improve representation learning and predictive accuracy. Neural Collaborative Filtering and transformer-based models have demonstrated improved performance in capturing nonlinear user–item interactions (Castells et al., 2021). However, many implementations in cosmetic retail continue to rely primarily on rating patterns or review sentiment without explicitly modeling the compatibility between individual skin characteristics and product ingredients (Chouhan & Barde, 2024). Skincare recommendations differ from general product recommendations because relevance depends on both preference similarity and biological suitability (Qalbyassalam et al., 2022). A product that performs well for oily skin may be unsuitable for dry or sensitive skin. Ingredient-level interactions play a central role in determining compatibility (Khan & Singh, 2025). Despite this, many existing neural recommender systems represent items using aggregated embeddings without examining fine-grained relationships between user attributes and ingredient composition (Gao & Meng, 2022). As a result, similarity is often computed through a static dot product, assuming uniform interaction strength across all features. Two-Tower neural architectures have gained popularity in large-scale retrieval systems because they encode users and items separately and enable efficient similarity computation (Roy & Hasan, 2023). This structure supports scalable recommendations in real-time environments. Nevertheless, conventional Two-Tower implementations typically employ simple similarity measures, limiting their ability to model structured interactions between user biological attributes and ingredient representations (Su et al., 2025). For skincare applications where compatibility may depend on specific ingredient–skin interactions, this simplification can reduce recommendation quality (Shah et al., 2022). This study proposes an Attention-Enhanced Neural Two-Tower architecture for personalized skincare product recommendation. The model integrates three primary information sources: (i) user behavioral interactions, (ii) explicit skin profile attributes, and (iii) ingredient-level textual metadata. A cross-attention mechanism is introduced to enable user embeddings to attend to ingredient representations prior to similarity scoring selectively. Cosine similarity is then applied within a shared embedding space to generate ranked recommendations. The architecture is trained using a contrastive objective with in-batch negative sampling. Experimental evaluation is conducted using the Amazon Beauty dataset compiled by (He & McAuley, 2024). Performance is assessed through standard ranking metrics, including Hit Ratio (HR@K), Normalized Discounted Cumulative Gain (nDCG@K), and Mean Reciprocal Rank (MRR). In addition to overall performance, results are analyzed across different skin types to examine variation in recommendation effectiveness. The main contributions of this

https://dx.doi.org/10.4314/swj.v21i1.46

study are summarized as follows: (1) Development of a neural retrieval architecture that integrates behavioral data, ingredient information, and explicit skin attributes within a unified embedding framework. (2) Introduction of cross-attention mechanism to model fine-grained interactions between user profiles and ingredient representations, and (3) Empirical evaluation of personalized skincare recommendation performance using ranking-based metrics and skin-type stratified analysis.

The remainder of this paper is organized as follows. Section 1.2 reviews related research in neural recommender systems and cosmetic recommendation approaches. Section 2 presents the proposed architecture and training procedure. Section 3 describes the experimental setup and results. Section 4 concludes the paper and outlines directions for further research.

Collaborative Filtering and Traditional Recommendation Approaches

Early developments in recommender systems for e-commerce relied heavily on collaborative filtering (CF) techniques, which infer user preferences based on historical interactions and similarities among users or items. Item-based and user-based CF models have been widely adopted due to their simplicity and effectiveness in capturing behavioral patterns (Han, 2020; Hidayat et al., 2025). In the skincare domain, CF has been applied to recommend products based on peer ratings, demonstrating reasonable predictive accuracy despite high sparsity in user–item interactions (Hidayat et al., 2025). However, traditional CF approaches exhibit several limitations. They are highly sensitive to data sparsity and cold-start problems, especially when new users or products lack sufficient interaction history (Nanthini & Kumar, 2023; Shah et al., 2022). Additionally, these models assume uniform rating behavior across users, which can introduce bias due to the subjectivity of rating scales. To address this issue, normalization-based CF methods have been proposed to standardize user ratings and improve fairness and predictive performance (Panda et al., 2020). Despite these improvements, CF-based systems remain fundamentally limited in domains such as skincare, where the quality of recommendations depends on both preference similarity and biological compatibility. These models typically fail to account for product composition or individual physiological differences, which results in recommendations that may be unsuitable for specific skin conditions (Chouhan & Barde, 2024; Khan & Singh, 2025).

Neural and Deep Learning-Based Recommender Systems

The limitations of traditional CF methods have led to the adoption of neural network-based approaches capable of modeling complex non-linear interactions between users and items. Neural Collaborative Filtering (NCF) has emerged as a foundational framework that replaces linear matrix factorization with deep neural architectures to improve representation learning (Shah et al., 2022; Manurung & Baizal, 2024). These models leverage implicit feedback, such as clicks and browsing behavior, to capture richer user preferences, thereby achieving significant improvements in recommendation accuracy under sparse conditions (Shah et al., 2022). Further advancements incorporate auxiliary data sources to enhance predictive performance. For instance, sentiment-aware NCF models leverage textual reviews to generate implicit ratings, resulting in substantial reductions in prediction error compared to traditional rating-based approaches (Qalbyassalam et al., 2022). Similarly, hybrid neural frameworks integrate multiple modalities such as reviews, ratings, and user feedback signals to provide more comprehensive preference modeling (Ibrahim et al., 2023). Recent developments also explore deep architectures that address data sparsity and multi-criteria decision-making. Autoencoder-based models combined with deep neural networks have demonstrated effectiveness in reconstructing missing ratings and capturing non-linear relationships in user–item interactions (Spoorthy et al., 2023). Transformer-based models and BERT-enhanced recommender systems further extend this capability by extracting contextual representations from textual data, enabling deeper semantic understanding of user preferences (Dinata & Baizal, 2024). While these approaches significantly improve predictive accuracy, many still focus on behavioral and textual signals without explicitly modeling domain-specific constraints, such as biological compatibility, in skincare applications.

Ingredient-Aware and Biologically Informed Recommendation Approaches

In cosmetics, the effectiveness of recommendations depends strongly on the compatibility between product ingredients and individual skin characteristics. As a result, several studies have emphasized the importance of incorporating ingredient-level information into recommendation systems. Ingredient-aware approaches treat product composition as a primary feature in determining suitability, enabling more targeted and safer recommendations (Honma et al., 2018; Chaurasia et al., 2022). One of the earliest contributions introduced an Ingredient Frequency–Inverse Product Frequency (IF-IPF) framework to identify ingredients that are most effective for specific user groups based on demographic and skin-type attributes (Honma et al., 2018). Similarly, content-based models have been used to map ingredient compositions to user skin profiles, achieving high accuracy in matching products to individual needs (Chaurasia et al., 2022). More recent research has explored deep learning techniques for modeling ingredient interactions. Transformer-based architectures have been applied to treat ingredient lists as sequential data to enable prediction of product efficacy across multiple dermatological concerns (Jinhee et al., 2024). In parallel, image-based approaches leverage computer vision models to analyze skin conditions directly and generate recommendations based on detected features such as acne, dryness, or pigmentation (Nguyen & Akter, 2025). Although these approaches introduce biological awareness into recommendation systems, many of them either treat ingredient information independently of user behavior or lack mechanisms to dynamically align user attributes with specific product components. This limits their ability to capture fine-grained compatibility relationships required for personalized skincare recommendations.

Hybrid and Multimodal Recommendation Systems

To overcome the limitations of single-method approaches, hybrid recommendation systems have been developed to combine multiple data sources and modeling techniques. These systems integrate collaborative filtering with content-based filtering, leveraging both user interaction data and product metadata to improve recommendation quality (Utami et al., 2025). Weighted hybrid models have shown that optimal performance depends on user characteristics such as tenure and interaction history. Content-based components tend to perform better for new users, while collaborative signals become more effective as more interaction data becomes available (Utami et al., 2025). Similarly,

https://dx.doi.org/10.4314/swj.v21i1.46

hybrid neural frameworks incorporate multiple modalities, including reviews, ratings, and sentiment information, to capture diverse aspects of user preferences (Ibrahim et al., 2023). Advancements in multimodal learning have further expanded the scope of recommendation systems. Models integrating textual, visual, and behavioral data have demonstrated improved performance in capturing complex user-item relationships (Dinata & Baizal, 2024; Jinhee et al., 2024). However, despite these improvements, most hybrid systems still lack explicit mechanisms for modeling biological compatibility, particularly in the context of skincare where ingredient–skin interactions are critical.

Two-Tower Architectures and Interaction Modeling

Two-Tower neural architectures have gained prominence in large-scale recommendation systems due to their scalability and efficiency. These models encode users and items independently into a shared embedding space, enabling fast similarity computation for real-time retrieval (Roy & Hasan, 2023; Su et al., 2025). Their modular structure makes them suitable for industrial applications with large datasets and low-latency requirements. Recent research has focused on enhancing interaction modeling within Two-Tower frameworks. Standard implementations rely on simple similarity measures such as dot product or cosine similarity, which assume uniform interaction strength across features. To address this limitation, advanced architectures introduce interaction modules that enable feature-level or representation-level alignment (Xiong et al., 2025). For example, fully interacted Two-Tower models incorporate mechanisms for early and late-stage feature interaction while maintaining computational efficiency (Xiong et al., 2025). Similarly, dual-tower models with hierarchical attention and sequence-aware sampling have demonstrated improved ability to capture temporal and demographic patterns in user behavior (Sundi, 2025). Despite these advancements, existing Two-Tower approaches generally do not incorporate biologically meaningful interactions between user attributes and product composition. This limitation is particularly significant in skincare recommendations, where relevance depends on condition-specific compatibility rather than general similarity.

Research Gap and Study Contribution

A review of existing literature indicates that significant progress has been made in recommender systems through collaborative filtering, deep learning, and hybrid modeling approaches (Dinata & Baizal, 2024; Hidayat et al., 2025; Qalbyassalam et al., 2022). However, several critical gaps remain. First, most models prioritize behavioral or textual signals without explicitly incorporating biological attributes such as skin type or sensitivity. Second, although ingredient-aware methods exist, they often lack dynamic interaction mechanisms that align user characteristics with specific product components. Third, conventional Two-Tower architectures do not adequately model fine-grained compatibility relationships required for domain-specific recommendations.

This study addresses these limitations by proposing a Multi-Head Attention Two-Tower architecture that integrates user behavior, explicit skin profiles, and ingredient-level representations within a unified framework. The introduction of attention-based interaction enables the model to dynamically evaluate compatibility between user attributes and product ingredients, thereby supporting biologically informed and personalized skincare recommendations.

MATERIALS AND METHODS

This study employs a quantitative experimental approach to develop and evaluate a personalized skincare recommendation system.

High Methodological Flow of the Study

The methodology is organized into four sequential stages: data acquisition and preprocessing, architecture design, interaction modeling, and model evaluation, as shown in Figure I.

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.46
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

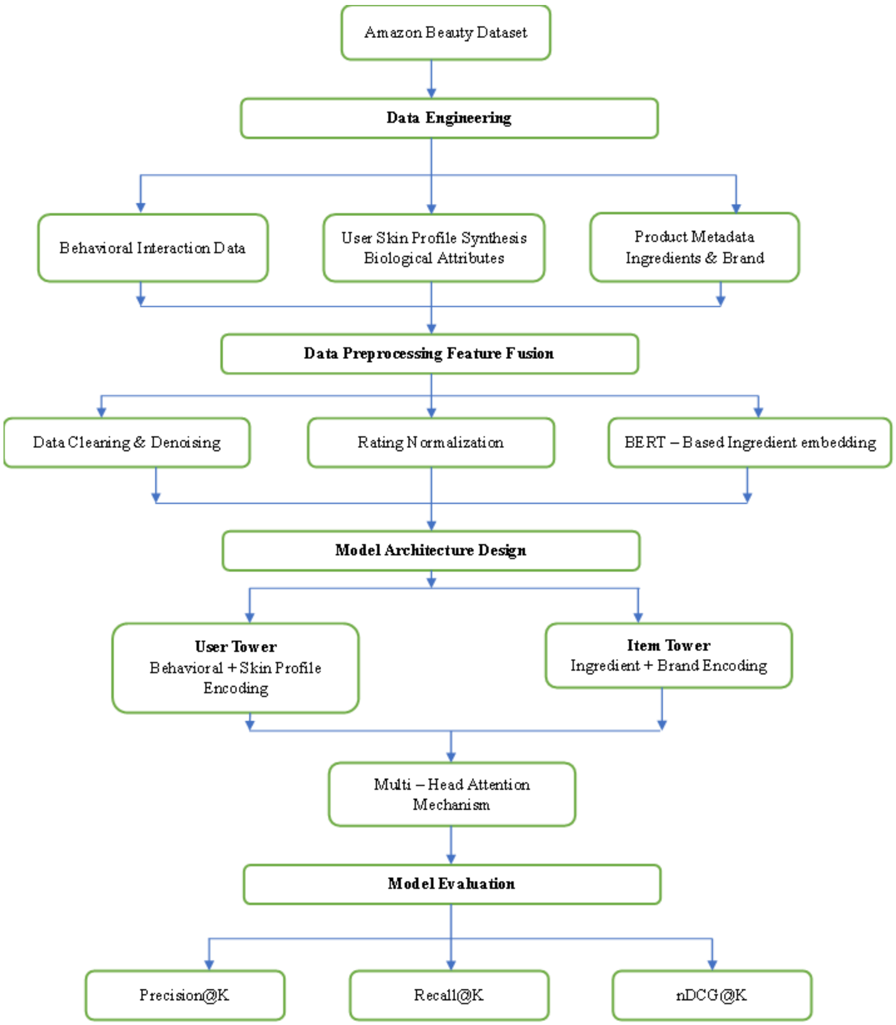


Figure 1 High Methodological Flow of the Study

Data Acquisition and Dataset Characteristics
To support reproducibility and empirical robustness, this study uses the expanded Amazon Beauty Dataset (He & McAuley, 2024), which is widely used in recommender system research. The dataset provides a rich combination of user interaction data and product metadata, which makes it well-suited for modeling personalized recommendations in the cosmetics domain.

Behavioral Data Source
The dataset comprises large-scale customer reviews and associated metadata collected from Amazon.com. For this research, emphasis is placed on the "All Beauty" and "Luxury Beauty" subcategories, as these segments contain detailed information on skincare and cosmetic products relevant to biological compatibility analysis.

Key characteristics of the dataset include:
Total Records: The full "All Beauty" subset contains approximately 2.02 million validated reviews and ratings.

- **User and Item Volume**: The dataset includes approximately 1.21 million unique users and 249,000 unique products identified through Amazon Standard Identification Numbers (ASINs).
- **Interaction Sparsity:** As with most large-scale e-commerce datasets, the user–item interaction matrix is highly sparse with most users interacting with only a small fraction of available products (He & McAuley, 2024).

To address sparsity and ensure stable model training, this study employs the 5-core filtered subset, which retains only users and items with at least 5 recorded interactions. This filtering process significantly increases data density and reduces noise, resulting in a compact yet informative dataset of approximately 5,269 interaction records suitable for deep neural training (Castells et al., 2021).

Feature Metadata and Ingredient Extraction
This research extends the dataset by extracting and enriching product- and user-level features beyond behavioral interactions. Unlike general-purpose recommender systems, skincare recommendations require explicit modeling of product composition and user biology.

Science World Journal Vol. 21(No 1) 2026
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v21i1.46

- **Item Features**
  Product metadata was extracted from Amazon's item metadata files and includes ASIN, brand name, product title, and detailed ingredient descriptions. Ingredient lists are standardized according to the International Nomenclature of Cosmetic Ingredients (INCI), where available, ensuring consistency in chemical representation.

- **User Features**
  In addition to basic identifiers such as reviewer ID and interaction timestamps, explicit skin profile attributes (e.g., oily, dry, sensitive, combination) are synthetically mapped to users. This process simulates real-world dermatological data-collection interfaces commonly found on modern beauty platforms (Utami et al., 2025).

Data Preprocessing and Feature Fusion
Prior to model training, the raw dataset underwent systematic preprocessing and feature fusion pipeline designed to enhance data quality and ensure compatibility with deep neural architectures. The pipeline consists of three primary stages:

1. **Data Cleaning and Denoising**
   Ingredient descriptions and product metadata were cleaned to remove HTML tags, special symbols, and non-ASCII characters. This step ensures textual consistency and reduces noise that may interfere with downstream embedding processes (Bin, 2023).
2. **Normalization of Behavioral Signals**
   User ratings, originally recorded on a five-point scale, were normalized to a continuous range between 0 and 1. This transformation mitigates the influence of individual rating habits and extreme scoring behaviors, which have been shown to bias learning in recommender systems (Panda et al., 2020). Normalization also facilitates stable gradient updates during neural network training.
3. **Textual Tokenization and Embedding**
   Product ingredient lists were processed using (Bidirectional Encoder Representations from Transformers) BERT-based tokenizer. Each ingredient sequence was converted into a dense 768-dimensional embedding to capture semantic relationships between chemical components. This approach enables the model to recognize functional similarities between ingredients even when they are expressed using different terminologies or formulations.

The Proposed Architecture: Multi-Head Attention Two-Tower Model
This study proposes a Multi-Head Attention-enhanced Two-Tower neural architecture designed to address the interaction limitations of conventional dual-encoder recommender systems in the cosmetics domain. Figure II illustrates that the architecture retains the computational efficiency and scalability of standard Two-Tower models while introducing a dynamic interaction mechanism that models biologically informed user–item relationships.

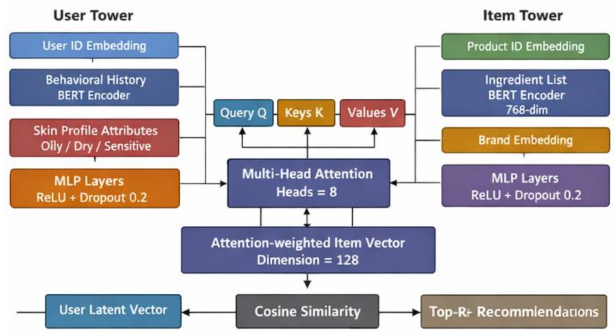


Figure 2 Architecture of the Proposed Model

At a high level, the model consists of two independent deep neural encoders: a User Tower and an Item Tower. Each tower transforms high-dimensional heterogeneous input features into dense representations within the shared latent space of dimension *d*. Unlike classical implementations that compute relevance through a static similarity function, the proposed architecture inserts a Multi-Head Attention (MHA) interaction layer between the towers. This layer enables model to dynamically evaluate how specific product ingredients align with individual user skin characteristics and allows relevance to be conditioned on biological compatibility rather than popularity alone.

Tower Encoders and Feature Embedding
Each tower in the proposed architecture is responsible for encoding a distinct feature space into a compact latent representation. The encoders are structurally symmetrical but semantically asymmetric to reflect the fundamental difference between user intent and product composition.

User Tower: Encoding Behavioral and Biological Context
The User Tower processes a composite feature vector that represents both implicit behavior and explicit biological attributes. Specifically, the user input includes:

- **User Identifier**: A categorical identifier that captures long-term preference signals.
- **Behavioral History**: Aggregated implicit interactions such as past purchases or engagements, representing habitual tendencies.
- **Skin Profile Attributes**: Explicit biological descriptors (e.g., oily, dry, sensitive) that impose compatibility constraints on recommendations.

These features were first mapped into dense vectors through embedding layers. Textual or sequential behavioral components will be encoded using a pretrained language representation model, while categorical attributes were embedded using learnable lookup tables. The resulting embeddings are concatenated and passed through a Multi-Layer Perceptron (MLP) that consists of fully connected layers with non-linear activation functions. This transformation produces final user latent vector, which summarizes both preference history and biological requirements.

Item Tower: Encoding Product Composition and Metadata
The Item Tower encodes product-level information, with an emphasis on chemical composition, which is central to skincare suitability. The item input includes:

- **Product Identifier:** Captures latent popularity and collaborative signals.

https://dx.doi.org/10.4314/swj.v21i1.46

- **Ingredient List:** A high-dimensional textual sequence describing the chemical formulation of the product.
- **Brand Information:** Encodes brand-level characteristics and formulation philosophies.

Ingredient lists were tokenized and embedded with a contextual language model, allowing semantically related ingredients to share representational proximity. Brand and identifier features were embedded separately and concatenated with ingredient representation. The combined vector was then processed through an MLP to generate the item latent representation in order to preserve both compositional detail and categorical context. At this stage, user and item representations are learned independently.

Mathematical Formulation of the Attention Mechanism

To overcome the interaction bottleneck in traditional Two-Tower models, this study introduces the Multi-Head Attention (MHA) mechanism as the core interaction layer. The attention module allows the model to selectively emphasize specific product ingredients based on user-specific biological signals rather than assuming fixed similarity between user and item embeddings.
In this formulation:

- The Query (Q) vector is derived from the final output of the User Tower and represents the user's current skin needs and preferences.
- The Keys (K) and Values (V) are derived from token-level representations of the product's ingredient list generated by Item Tower.

For each attention head, the user representation queries ingredient tokens to compute relevance scores, enabling the model to focus on ingredients most compatible with the user's skin profile. Multiple attention heads are employed to capture different perspectives on compatibility, such as hydration, irritation potential, and oil control. The scaled dot-product attention for a single head and aggregation of multiple heads is defined in equation 1 as follows:

$$head_i = Attention\left(QW^Q{}_i, KW^K{}_i, VW^V{}_i,\right) = Softmax\left(\frac{\left(QW^Q{}_i\right)\left(KW^K{}_i\right)^T}{\sqrt{d_k}}\right) \quad (1)$$

here $W^Q{}_i, W^K{}_i, W^V{}_i \in R^{d \times d_k}$ are learnable projection matrices and d_k is the dimensionality of the key vectors used for scaling (Su et al., 2025).
The outputs of all attention heads are concatenated and projected through a final linear transformation: (see equation 2)

$$MHA(Q,K,V) = Concat(head_i, \ldots \ldots head_h)W^o \quad (2)$$

This process yields an attention-weighted item representation that is explicitly conditioned on the user's biological context (Roy & Hasan, 2023).

Interaction and Similarity Scoring

Following attention-based fusion, the model produces two aligned vectors:

- a user latent vector, representing preferences and skin constraints, and
- an attention-refined item vector, representing the subset of ingredients most relevant to that user.

The final relevance score determines ranking during recommendation. To ensure stability and scale invariance, similarity is computed using cosine similarity rather than a raw dot product. This choice normalizes vector magnitudes and focuses the scoring on directional alignment within the latent space.

The similarity function is defined in equation 3 as:

$$S(u,v) = \cos(u,v') = \frac{u \,.\, v'}{||u||v'||} \quad (3)$$

This scoring strategy preserves the retrieval efficiency of Two-Tower architectures while enabling recommendations that are dynamically informed by ingredient-level compatibility. As a result, the proposed model delivered recommendations that are both personalized and biologically appropriate to address the core challenges identified in the cosmetics recommendation domain (Ibrahim et al., 2023).

Model Training and Parameters

Training was conducted using historical interaction data. During each training step, the model is exposed to batches of user–item interactions, enabling it to learn discriminative representations that distinguish suitable products from less appropriate alternatives. This approach aligns with the study's objective of producing safe and personalized skincare recommendations.

Data Splitting Strategy

- **Training Set** (70%) – was used to train the models and learn embedding weights.
- **Validation Set** (15%) – was used for hyperparameter tuning (e.g., learning rate, embedding dimension, regularization weight).
- **Test Set** (15%) – was a held-out set used exclusively for final performance evaluation.

Training Hyperparameters

The experimental setup uses the following key hyperparameters:

- **Embedding Dimension:** 128, providing a balance between expressive capacity and efficiency.
- **Attention Heads:** 8, allowing the model to capture multiple compatibility patterns simultaneously.
- **Optimizer and Learning Rate**: Adam optimizer with a learning rate of 1× 10^{-4}
- **Regularization:** Dropout rate of 0.2 applied to fully connected layers to reduce overfitting.
- **Training epochs:** 60, with early stopping set at 20 epochs if the models learned all necessary parameters.

Evaluation Metrics

Given the ranking-oriented nature of recommender systems, classification accuracy or Mean Absolute Error (MAE) is an insufficient indicator of performance. Instead, this study employs three widely recognized ranking metrics: Precision@K, Recall@K, and Normalized Discounted Cumulative Gain (nDCG@K). Each metric captures a distinct dimension of recommendation quality, such as accuracy, completeness, and rank sensitivity.
All metrics were computed at K = 10 and K = 50 to evaluate both short-list (high precision) and broad-list (high recall) recommendation quality.

(1) Precision@K

Precision@K measures the proportion of correctly recommended (relevant) items among the top-K list generated for each user. It evaluates the precision of the system's top-ranked suggestions. (see equation 4)

$$\text{Precision@K} = \frac{|R_u^K \cap T_u|}{K} \quad (4)$$

Where R_u^K = the set of top-K items recommended to user u; T_u =

https://dx.doi.org/10.4314/swj.v21i1.46

the set of ground-truth relevant items for user u; $|\cdot|$ denotes set cardinality. A higher Precision@K value indicates that the top recommendations align closely with the user's actual preferences (Gao & Li, 2022).

(2) Recall@K
Recall@K measures the proportion of relevant items that appear in the top-K recommendations. Unlike precision, which rewards focus, recall emphasizes completeness, such as how many of the items the user actually liked were retrieved. (see equation 5)

$$\text{Recall@K} = \frac{|R_u^K \cap T_u|}{T_u} \qquad (5)$$

Where T_u= total number of relevant items for user u. High Recall@K values indicate that the system successfully identifies most of the items a user would find relevant, even if not all are ranked near the top (Utami et al., 2025).

(3) Normalized Discounted Cumulative Gain (nDCG@K)
nDCG@K is considered the gold-standard metric for ranking evaluation, as it incorporates both the relevance and the position of items in the recommendation list. It assigns greater weight to relevant items that appear earlier in the ranking. (see equation 6)

$$\text{nDCG@K} = \frac{1}{\text{IDCG@K}} \sum_{i=1}^{K} \frac{2^{reli} - 1}{\log_2(i+1)} \qquad (6)$$

Where $reli$ = relevance score (1 if item i is relevant, 0 otherwise); i = rank position in the top-K list; IDCG@K = the ideal (maximum possible) DCG value at position K used to normalize results into the range [0,1]. A higher nDCG@K indicates that relevant items are retrieved and ranked higher in the recommendation list, reflecting a better real-world user experience (Ibrahim et al., 2023).

System Implementation and Simulation Environment
The implementation of the proposed Multi-Head Attention Two-Tower model requires a high-performance computing environment to handle the parallel processing of dense embeddings and the training of the transformer-based attention layers.

Hardware Specifications
To ensure efficient training and low-latency inference during evaluation, the following hardware configuration is utilized:

- **Processor (CPU):** Intel Core i7 to manage data pre-processing and multi-threaded data loading.
- Memory (RAM): 32GB DDR4 3200MHz to facilitate handling of large-scale behavioral datasets and ingredient metadata in-memory.

Software and Development Tools
The system will be developed using Python (v3.9) programming language, chosen for its extensive ecosystem of machine learning libraries.

- **Deep Learning Framework**: PyTorch 2.0 or TensorFlow 2.10, utilized for constructing the dual-tower architecture and implementing the custom attention layers.
- **Natural Language Processing**: HuggingFace Transformers library is used to load the pre-trained BERT weights for ingredient tokenization.
- **Data Manipulation**: Pandas and NumPy for data cleaning and matrix operations.
- **Evaluation:** Scikit-learn and custom ranking scripts to calculate Recall@K and nDCG@K.

RESULTS AND DISCUSSION

Training Performance and Convergence
The proposed Multi-Head Attention Two-Tower model demonstrated rapid convergence during training. Using Triplet Margin Loss with hard negative sampling, training loss decreased from 0.3868 in the first epoch to 0.0004 by epoch 2, while validation loss remained stable throughout 20 epochs (Figures II and III). Minimal divergence between training and validation losses indicates balanced model expressivity and generalization, suggesting effective learning despite the dataset's sparsity (Nanthini & Kumar, 2023).

Science World Journal Vol. 21(No 1) 2026
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v21i1.46

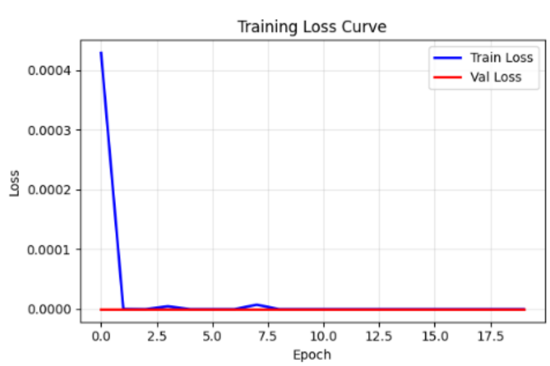
```
1. Creating datasets with negative samples...
Training with 4320 batches (with negatives)

2. Initializing fresh model...
3. Starting training with negative sampling...
Epoch 1/20
  Batch 0/4320, Loss: 0.3868
  Batch 50/4320, Loss: 0.0000
  Batch 100/4320, Loss: 0.0000
  Batch 150/4320, Loss: 0.0000
  Batch 200/4320, Loss: 0.0000
  Batch 250/4320, Loss: 0.0000
  Batch 300/4320, Loss: 0.0039
  Batch 350/4320, Loss: 0.0000
  Batch 400/4320, Loss: 0.0000
  Batch 450/4320, Loss: 0.0000
  Batch 500/4320, Loss: 0.0000
  Batch 550/4320, Loss: 0.0000
  Batch 600/4320, Loss: 0.0000
  Batch 650/4320, Loss: 0.0000
  Batch 700/4320, Loss: 0.0000
  Batch 750/4320, Loss: 0.0000
  Batch 800/4320, Loss: 0.0000
  Batch 850/4320, Loss: 0.0000
  Batch 900/4320, Loss: 0.0000
  Batch 950/4320, Loss: 0.0000
  Batch 1000/4320, Loss: 0.0000
  Batch 1050/4320, Loss: 0.0000
  Batch 1100/4320, Loss: 0.0000
  Batch 1150/4320, Loss: 0.0000
  Batch 1200/4320, Loss: 0.0000
  Batch 1250/4320, Loss: 0.0000
  Batch 1300/4320, Loss: 0.0000
  Batch 1350/4320, Loss: 0.0000
  Batch 1400/4320, Loss: 0.0000
  Batch 1450/4320, Loss: 0.0000
  Batch 1500/4320, Loss: 0.0000
  Batch 1550/4320, Loss: 0.0000
  Batch 1600/4320, Loss: 0.0000
  Batch 1650/4320, Loss: 0.0000
  Batch 1700/4320, Loss: 0.0000
```

```
Epoch 2/20
  Batch 0/4320, Loss: 0.0000
  Batch 50/4320, Loss: 0.0000
  Batch 100/4320, Loss: 0.0000
  Batch 150/4320, Loss: 0.0000
  Batch 200/4320, Loss: 0.0000
  Batch 250/4320, Loss: 0.0000
  Batch 300/4320, Loss: 0.0000
  Batch 350/4320, Loss: 0.0000
  Batch 400/4320, Loss: 0.0000
  Batch 450/4320, Loss: 0.0000
  Batch 500/4320, Loss: 0.0000
  Batch 550/4320, Loss: 0.0000
  Batch 600/4320, Loss: 0.0000
  Batch 650/4320, Loss: 0.0000
  Batch 700/4320, Loss: 0.0000
  Batch 750/4320, Loss: 0.0000
  Batch 800/4320, Loss: 0.0000
  Batch 850/4320, Loss: 0.0000
  Batch 900/4320, Loss: 0.0000
  Batch 950/4320, Loss: 0.0000
  Batch 1000/4320, Loss: 0.0000
  Batch 1050/4320, Loss: 0.0000
  Batch 1100/4320, Loss: 0.0000
  Batch 1150/4320, Loss: 0.0000
  Batch 1200/4320, Loss: 0.0000
  Batch 1250/4320, Loss: 0.0000
  Batch 1300/4320, Loss: 0.0000
  Batch 1350/4320, Loss: 0.0000
  Batch 1400/4320, Loss: 0.0000
  Batch 1450/4320, Loss: 0.0000
  Batch 1500/4320, Loss: 0.0000
  Batch 1550/4320, Loss: 0.0000
  Batch 1600/4320, Loss: 0.0000
  Batch 1650/4320, Loss: 0.0000
  Batch 1700/4320, Loss: 0.0000
  Batch 1750/4320, Loss: 0.0000
  Batch 1800/4320, Loss: 0.0000
  Batch 1850/4320, Loss: 0.0000
  Batch 1900/4320, Loss: 0.0000
  Batch 1950/4320, Loss: 0.0000
  Batch 2000/4320, Loss: 0.0000
```

```
Epoch 20/20
  Batch 0/4320, Loss: 0.0000
  Batch 50/4320, Loss: 0.0000
  Batch 100/4320, Loss: 0.0000
  Batch 150/4320, Loss: 0.0000
  Batch 200/4320, Loss: 0.0000
  Batch 250/4320, Loss: 0.0000
  Batch 300/4320, Loss: 0.0000
  Batch 350/4320, Loss: 0.0000
  Batch 400/4320, Loss: 0.0000
  Batch 450/4320, Loss: 0.0000
  Batch 500/4320, Loss: 0.0000
  Batch 550/4320, Loss: 0.0000
  Batch 600/4320, Loss: 0.0000
  Batch 650/4320, Loss: 0.0000
  Batch 700/4320, Loss: 0.0000
  Batch 750/4320, Loss: 0.0000
  Batch 800/4320, Loss: 0.0000
  Batch 850/4320, Loss: 0.0000
  Batch 900/4320, Loss: 0.0000
  Batch 950/4320, Loss: 0.0000
  Batch 1000/4320, Loss: 0.0000
  Batch 1050/4320, Loss: 0.0000
  Batch 1100/4320, Loss: 0.0000
  Batch 1150/4320, Loss: 0.0000
  Batch 1200/4320, Loss: 0.0000
  Batch 1250/4320, Loss: 0.0000
  Batch 1300/4320, Loss: 0.0000
  Batch 1350/4320, Loss: 0.0000
  Batch 1400/4320, Loss: 0.0000
  Batch 1450/4320, Loss: 0.0000
  Batch 1500/4320, Loss: 0.0000
  Batch 1550/4320, Loss: 0.0000
  Batch 1600/4320, Loss: 0.0000
  Batch 1650/4320, Loss: 0.0000
  Batch 1700/4320, Loss: 0.0000
  Batch 1750/4320, Loss: 0.0000
  Batch 1800/4320, Loss: 0.0000
  Batch 1850/4320, Loss: 0.0000
  Batch 1900/4320, Loss: 0.0000
  Batch 1950/4320, Loss: 0.0000
  Batch 2000/4320, Loss: 0.0000
```

Figure 3 Training Process of the Model

This accelerated convergence indicates effective gradient propagation through the dual-tower architecture and validates the stability of the attention-based interaction layer.




```
Final Train Loss: 0.000000
Final Val Loss: 0.000000
```

+ Code + Markdown

Figure 4. Training Loss Curve

Ranking Performance

Recommendation performance was assessed using leave-one-out evaluation with 100 candidate items per user. The model achieved a Hit Rate of 12.41% at K = 10, which is significantly higher than random selection (≈1%), confirming meaningful learning from behavioral and ingredient-level signals as shown in Figure V. (Chaurasia et al., 2022).

```
Running comprehensive evaluation...

==================================================
EVALUATION RESULTS (200 users)
==================================================

K = 1:
  Hit Rate@1:      0.0073
  Precision@1:     0.0073
  Recall@1:        0.0012
  NDCG@1:          0.0073

K = 5:
  Hit Rate@5:      0.0584
  Precision@5:     0.0117
  Recall@5:        0.0185
  NDCG@5:          0.0342

K = 10:
  Hit Rate@10:      0.1241
  Precision@10:     0.0124
  Recall@10:        0.0395
  NDCG@10:          0.0554

K = 20:
  Hit Rate@20:      0.1971
  Precision@20:     0.0099
  Recall@20:        0.0664
  NDCG@20:          0.0737

Metrics saved to /kaggle/working/evaluation_metrics.csv
```

Figure 5 Primary Ranking Performance

Table 1 summarizes ranking performance across multiple cut-off points. Hit Rate increased from 0.73% at K = 1 to 19.71% at K = 20, while Precision peaked at K = 10 (1.24%). Recall and nDCG showed consistent improvement as K increased, indicating effective retrieval of relevant items. These results demonstrate that

Science World Journal Vol. 21(No 1) 2026
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v21i1.46

integrating attention-based ingredient representations alongside behavioral history enhances recommendation quality in sparse data conditions (Qalbyassalam et al., 2022; Utami et al., 2025).

Table 1: Ranking Performance Metrics at Various Cut-off Points

Metric	**K=1**	**K=5**	**K=10**	**K=20**	**Interpretation**
Hit Rate	0.73%	5.84%	12.41%	19.71%	Probability of a relevant item appearing in top-K
Precision	0.73%	1.17%	1.24%	0.99%	Proportion of relevant items in top-K
Recall	0.12%	1.85%	3.95%	6.64%	Proportion of relevant items retrieved
nDCG	0.73%	3.42%	5.54%	7.37%	Position-weighted relevance measure

Rank Distribution Analysis
While aggregate ranking metrics provide an overall assessment of model performance, rank distribution analysis offers deeper insight into how consistently the model positions relevant items across users. Figures VI and VII illustrate the distribution of ranks assigned to the held-out items, revealing a long-tailed pattern with meaningful concentration at lower ranks.

```
==================================================
RANK STATISTICS
==================================================
Total users evaluated: 137
Mean rank: 49.0
Median rank: 48.0
Std rank: 28.2
Min rank: 1
Max rank: 100

Percentiles:
  10th percentile: 8.0
  25th percentile: 29.0
  50th percentile: 48.0
  75th percentile: 72.0
  90th percentile: 86.0
  95th percentile: 94.2
```

+ Code   + Markdown

Figure 6 Key Distribution Analysis

The mean rank of 49.0 and median rank of 48.0 indicate that, on average, relevant items are placed near the midpoint of the candidate list. Given the challenging evaluation setup involving 100 candidate items per user, this result reflects performance that exceeds the random expectation, where the theoretical mean rank would be approximately 50.5 (Chouhan & Barde, 2024). The standard deviation of 28.2 highlights substantial variability in ranking quality across users, suggesting that differences influence the effectiveness of recommendations in terms of interaction density, behavioral signal strength, and the complexity of individual skin-profile-to-ingredient relationships (Nanthini & Kumar, 2023).

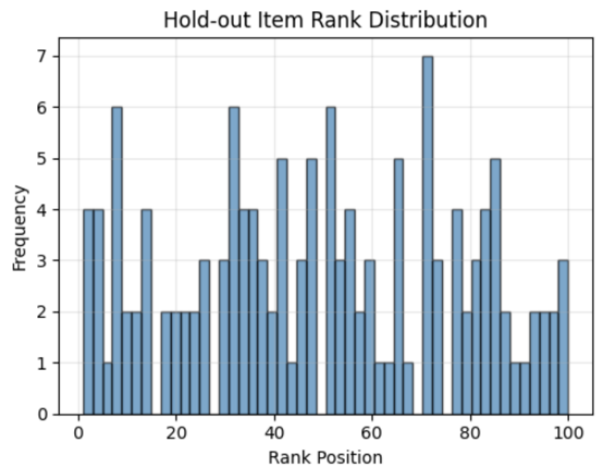


Figure 7 Held-out Item Rank Distribution

The cumulative rank distribution in Figure 6 further clarifies this variation. As reported in the percentile analysis, 10% of relevant items appear in the top 8 positions, and 25% in the top 29 positions. This indicates that a nontrivial subset of users benefits from highly accurate recommendations, with relevant products surfacing early in the ranking. At the same time, the presence of relevant items across full rank spectrum, which includes higher positions beyond rank 50 reflects difficulty of modeling personalized skincare suitability under sparse data conditions.

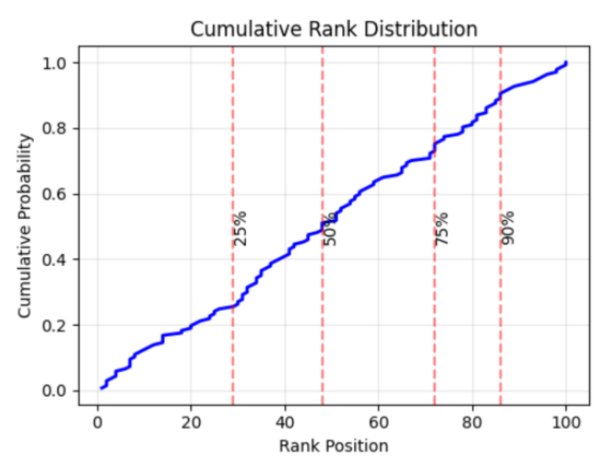


Figure 8 Cumulative Rank Distributions

The rank distribution analysis confirms that the model does not rely solely on popularity bias but instead produces differentiated ranking outcomes across users. This behavior is consistent with systems that learn user-specific compatibility patterns rather than uniform similarity functions, which support the study's emphasis on biologically informed recommendation modeling (S. Gao & Meng, 2022).

Biological Compatibility Analysis
A key outcome of this study is the observed variation in recommendation performance across different skin types, as illustrated in Figure IX. This stratified analysis provides evidence that the proposed model has learned biologically relevant compatibility patterns rather than relying solely on aggregate popularity signals.

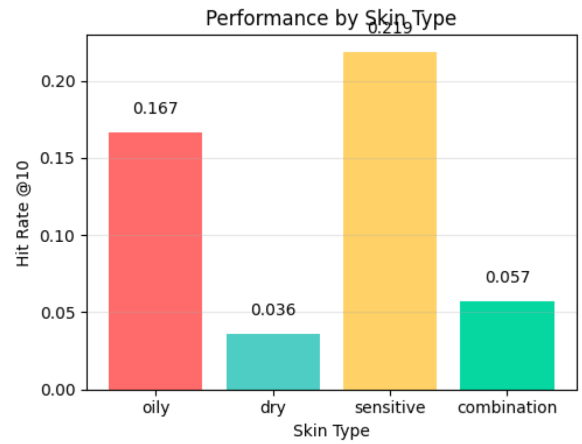


Figure 9 Performance by Skin Type

As summarized in Table II, the model achieved its strongest performance for users with sensitive skin, recording a Hit Rate of 21.9% at K = 10 and the lowest average rank (44.7). Sensitive skin presents the most restrictive compatibility constraints as products must avoid common irritants such as fragrances, alcohols, and certain preservatives. The model's superior performance for this group suggests effective learning of ingredient-level contraindications and highlights the contribution of the attention-based interaction mechanism in prioritizing safety-critical cases (Jinhee et al., 2024).

Table 2: Performance Stratified by Skin Type

Skin Type	Hit Rate @10	Average Rank	Relative Performance	Interpretation
Sensitive	21.9%	44.7	Best Performing	Model excels at identifying non-irritating formulations.
Oily	16.7%	47.0	Strong	Effectively matches oil-control ingredients.
Combination	5.7%	48.9	Moderate	Handles mixed skin characteristics
Dry	3.6%	57.1	Weakest	Challenges with hydration matching

Users with oily skin also showed strong performance, with a Hit Rate of 16.7% and an average rank of 47.0. This indicates that the model successfully aligns oil-control ingredients, such as salicylic acid and niacinamide, with their corresponding user profiles. Performance for combination skin was moderate, which reflects the complexity of mixed skin characteristics that require balancing competing formulation properties within a single recommendation (Qalbyassalam et al., 2022). The weakest performance was observed for users with dry skin, with a Hit Rate of 3.6% and a higher average rank of 57.1. This outcome suggests greater difficulty in modeling hydration-specific needs, potentially due to variability in ingredient effectiveness across users and the nature of dry-skin indicators in behavioral data. Similar challenges have been reported in prior studies on skincare recommendations, where hydration matching is more sensitive to individual skin and environmental factors (Manurung & Baizal, 2024). These performance disparities across skin types reinforce the argument that skincare recommendations are condition-dependent. The model's ability to deliver its strongest results for the most biologically constrained user group demonstrates the practical value of integrating explicit skin profiles and attention-based ingredient weighting within the Two-Tower architecture.

Analysis of Results
This section synthesizes the quantitative and qualitative findings presented in Sections 4.1 to 4.4, with emphasis on ranking

https://dx.doi.org/10.4314/swj.v21i1.46

effectiveness, personalization behavior, and biological relevance. The proposed model outperforms random baselines in highly constrained recommendation settings across all evaluation metrics. The consistent improvements in Hit Rate, Recall, and nDCG confirm that the architecture is effective at retrieving and prioritizing relevant skincare products, even with limited interaction history. These results suggest that the integration of ingredient-level representations meaningfully contributes to ranking quality.
The rank distribution and skin-type-specific analyses reveal that recommendation performance is not uniform across users. Stronger results for sensitive and oily skin highlight the model's ability to exploit explicit compatibility constraints, while weaker performance for dry skin highlights challenges associated with modeling diffuse and context-dependent needs. This variation reinforces the importance of condition-aware modeling in skincare recommendation systems. These findings demonstrate that biologically informed attention mechanisms enhance both effectiveness and interpretability of recommendation outcomes. The results support the central premise of this study that combining behavioral data with structured ingredient knowledge enables more reliable and condition-sensitive personalization in skincare recommendation systems.

Comparison with Related Works
To assess the performance of the proposed Multi-Head Attention Two-Tower model, it is instructive to compare it with related approaches in the skincare and broader recommender-system literature. While previous studies have employed collaborative filtering, hybrid models, or text-augmented neural networks (Dinata & Baizal, 2024; Honma et al., 2018; Qalbyassalam et al., 2022), few have integrated explicit biological attributes, such as skin type, with attention-based mechanisms to inform recommendations. Table III summarizes key distinctions across model architecture, data modalities, biological awareness, and reported performance metrics, highlighting the unique contribution of the present work in combining ingredient-level compatibility with user-specific skin profiles.

Table 3 Comparison with Related Works

Study	**Domain**	**Model / Approach**	**Data Modalities**	**Biological Awareness**	**Key Metrics / Findings**
(Honma et al., 2018)	Cosmetics	IF-IPF ingredient-aware recommender	User reviews + ingredient lists	X (implicit only)	Reduced invalid recommendations (<5%)
(Qalbyassalam et al., 2022)	Skincare	NCF with sentiment as implicit rating	Ratings + review text	X	RMSE improved from 0.8033 to 0.4931
(Dinata & Baizal, 2024)	Skincare	BERT-RS hybrid	Ratings + BERT review embeddings	Partial (explicit attrs included)	RMSE: 0.3039
(Utami et al., 2025)	Beauty (makeup)	Weighted CF+CBF	Ratings + text metadata	X	Precision ~83–90% (user tenure impact)
(Jinhee et al., 2024)	Skincare	Image + Transformer for ingredients	Images + ingredient text	Partial (via skin analysis)	Accuracy ~87–88%
(Su et al., 2025)	Recommender (general)	Two-Tower with multi-source fusion	Behavioral + metadata	X	AUC improvements in advertising
This study	Skincare	Attention-enhanced Two-Tower	Behavioral + text ingredients + explicit skin profile	✓	HR@10: 12.41%, meaningful compatibility patterns

Practical Implications and Recommendations
The findings provide several insights for the design and deployment of personalized skincare recommendation systems. First, the results demonstrate the importance of integrating biological characteristics into recommendation models. Systems that rely solely on interaction history or popularity signals may overlook ingredient safety and compatibility. Incorporating explicit skin profiles allows recommendation engines to prioritize products that align with individual physiological requirements. Second, the use of attention mechanisms offers practical advantage for ingredient-level reasoning. The Multi-Head Attention layer enables the model to highlight ingredients that contribute most strongly to compatibility with specific skin types. This capability improves transparency and can support explainable recommendation interfaces where users are informed why a particular product is suggested. Providing such explanations may strengthen user trust, especially in applications involving personal care and dermatological products. Third, the evaluation results suggest that biologically constrained conditions, such as sensitive skin, benefit significantly from compatibility-aware modeling. Developers of beauty and wellness platforms may therefore consider incorporating structured knowledge of ingredients and user skin attributes into their recommendation pipelines. This approach can reduce the likelihood of unsuitable product suggestions and improve the relevance of recommendations. Finally, practitioners should adopt training strategies that mitigate the effects of sparse interaction data. Techniques such as 5-core filtering, negative sampling, and representation learning help models learn meaningful patterns even when user feedback is limited. These strategies are especially relevant in specialized recommendation tasks where large-scale interaction data may not always be available.

https://dx.doi.org/10.4314/swj.v21i1.46

Limitations and Future Research

Although the proposed model demonstrates promising performance, several limitations should be acknowledged. The study relies on the Amazon Beauty dataset with 5-core filtering, which reduces sparsity but also limits the number of available interactions. As a result, the dataset may not fully capture the diversity of skincare routines, ingredient preferences, and product usage patterns observed in real-world settings. Another limitation concerns the representation of biological attributes. Performance differences across skin types also indicate areas for further improvement. The model performs well on sensitive and oily skin but shows weaker results on dry skin. Hydration-related skincare needs are influenced by multiple factors, including climate, lifestyle, and individual physiology. These aspects are not captured in the current dataset and may require richer feature representations.

Future research may address these limitations through several directions. Larger and more diverse datasets could improve generalization and allow models to learn broader compatibility patterns. Integration of additional biological signals, such as dermatological assessments or user-reported skin conditions, may further refine the accuracy of recommendations. Researchers may also explore alternative architectures, such as graph-based ingredient modeling or advanced attention mechanisms, to capture complex relationships among product components. Finally, real-world evaluation involving user feedback would provide valuable evidence regarding the effectiveness and usability of biologically informed skincare recommendation systems.

Conclusion

This study presented a personalized skincare recommendation model based on a Multi-Head Attention Two-Tower neural architecture. The model integrates behavioral interaction data, ingredient-level representations, and explicit skin profiles to generate compatibility-aware recommendations. Experimental evaluation shows that the model achieves a Hit Rate of 12.41% at K = 10 under a leave-one-out ranking protocol. Rank distribution analysis indicates that relevant products are frequently positioned near the midpoint of the candidate list, with a notable portion appearing in top-ranked positions. Performance differences across skin types further demonstrate the model's ability to capture biologically meaningful relationships between user characteristics and product ingredients. The results suggest that incorporating structured biological information improves the relevance of recommendations in skincare applications. Attention-based interaction modeling allows the system to emphasize ingredients that align with specific user profiles, which contributes to both recommendation quality and interpretability. The study highlights the potential of combining neural recommendation architectures with biologically informed features to support safer, more personalized discovery of skincare products. These findings provide a foundation for future research aimed at refining compatibility modeling and extending recommendation techniques to other health-related consumer applications.

REFERENCES

Baba, N., Sei, Y., Tahara, Y., & Ohsuga, A. (2024). Proposal of a Cosmetic Product Recommendation Method with Review Text that is Predicted to Be Written by Users. *ICAART (3)*, 609–616.

Bin, S. (2023). An e-commerce personalized recommendation algorithm based on multiple social relationships. *Sustainability*, *16*(1), 362.

Castells, P., Hurley, N., & Vargas, S. (2021). Novelty and diversity in recommender systems. In *Recommender Systems Handbook* (pp. 603–646). Springer.

Chaurasia, M., Pathak, N., Rani, M., Verma, M., & Gauhri, N. (2022). A machine learning based recommendation system for cosmetics. *Journal of Pharmaceutical Negative Results*, 3711–3716.

Chouhan, R., & Barde, S. (2024). A Review on Cosmetic Product Recommendation Using Deep Learning. *2024 4th International Conference on Soft Computing for Security Applications (ICSCSA)*, 1–8.

Dinata, A. A., & Baizal, Z. (2024). Rating prediction from user feedback reviews using BERT-RS. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1–5.

Gao, L., & Li, J. (2022). E-Commerce Personalized Recommendation Model Based on Semantic Sentiment. *Mobile Information Systems*, *2022*(1), 7246802.

Gao, S., & Meng, W. (2022). Development of a Personalized Recommendation System for E-Commerce Products for Distributed Storage Systems. *Computational Intelligence and Neuroscience*, *2022*(1), 4752981.

Han, K. (2020). Personalized news recommendation and simulation based on an improved collaborative filtering algorithm. *Complexity*, *2020*(1), 8834908.

He, R., & McAuley, J. (2024). *Amazon Beauty Dataset*. TIB/Github. https://service.tib.eu/ldmservice/dataset/amazon-beauty-dataset

Hidayat, A. A., Maulindar, J., & Indah, R. P. (2025). Application of Item-Based Collaborative Filtering Method for Skincare Recommendation System. *Journal of Advances in Information and Industrial Technology*, *7*(1), 13–20.

Honma, H., Nakajima, Y., & Aoshima, H. (2018). *Recommender system for cosmetics based on user evaluation and ingredient information*. University of Copenhagen, Denmark.

Ibrahim, M., Bajwa, I. S., Sarwar, N., Hajjej, F., & Sakr, H. A. (2023). An intelligent hybrid neural collaborative filtering approach for true recommendations. *IEEE Access*, *11*, 64831–64849.

Jinhee, L., Yoon, H., Kim, S., Lee, C., Lee, J., & Yoo, S. (2024). Deep learning-based skin care product recommendation: A focus on cosmetic ingredient analysis and facial skin conditions. *Journal of Cosmetic Dermatology*, *23*(6), 2066–2077. https://doi.org/10.1111/jocd.16218

Khan, R. E. H., & Singh, K. (2025). AI-Driven Personalized Skincare: Enhancing Skin Analysis and Product Recommendation Systems. *Journal of Scientific Innovation and Advanced Research (JSIAR)*, *1*(2), 178–184.

Manurung, T. A., & Baizal, Z. (2024). The Usage of Neural Collaborative Filtering in Enhancing Personalized Skincare Recommendation. *2024 International Conference on Data Science and Its Applications (ICoDSA)*, 375–379.

Nanthini, M., & Kumar, K. P. M. (2023). Cold Start and Data Sparsity Problems in Recommender Systems: A Concise Review. In D. Gupta, A. Khanna, S.

Bhattacharyya, A. E. Hassanien, S. Anand, & A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications* (pp. 107–118). Springer Nature Singapore.

Nguyen, A. T. P., & Akter, A. (2025). Intelligent Skincare: AI-Driven Cosmetic Product Recommendation Using Advanced Machine Learning Models. *American Journal of Health, Medicine and Nursing Practice*, *11*(1), 26–35.

Panda, S. K., Bhoi, S. K., & Singh, M. (2020). A collaborative filtering recommendation algorithm based on a normalization approach. *Journal of Ambient Intelligence and Humanized Computing*, *11*(11), 4643–4665.

Qalbyassalam, C., Rachmadi, R. F., & Kurniawan, A. (2022). Skincare recommender system using neural collaborative filtering with implicit rating. *2022 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, 272–277.

Roy, F., & Hasan, M. (2023). An item–item collaborative filtering recommender system based on item reviews: An approach with deep learning. *Vietnam Journal of Computer Science*, *10*(04), 517–536.

Shah, V., Anunay, & Kumar, P. (2022). Recommendation system using neural collaborative filtering and deep learning. *The International Conference on Recent Innovations in Computing*, 109–120.

Spoorthy, G., Sanjeevi, S., & others. (2023). Multi-criteria–recommendations using autoencoder and deep neural networks with weight optimization using the firefly algorithm. *International Journal of Engineering*, *36*(1), 130–138.

Su, Y., Li, Y., & Zhang, Z. (2025). Two-Tower Structure Recommendation Method Fusing Multi-Source Data. *Electronics*, *14*(5), 1-23 https://doi.org/10.3390/electronics14051003

Sundi, J. (2025). Personalised Product Discovery in Online Beauty Retail Using Dual-Tower Neural Networks. *Sarcouncil Journal of Engineering and Computer Sciences*, *4*(8), 437–446. https://doi.org/10.5281/zenodo.16814206

Utami, P. M. S., Trisna, I. N. P., & Vihikan, W. O. (2025). Web-Based Makeup Recommendation System Using Hybrid Filtering. *Journal of Applied Informatics and Computing*, *9*(3), 683–692.

Xiong, C., Yu, X., Xu, W., Cheng, L., Yuan, C., & Mo, L. (2025). A Learnable Fully Interacted Two-Tower Model for Pre-Ranking System. *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '25*, 2182–2191. https://doi.org/10.1145/3726302.3729881.