International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4958-4964
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0651

Ethnic Fit: Siamese Network-Based Outfit Compatibility Prediction for Indian Ethnic Wear

Mohamed Moyiz Khan[1], Dr. Manjula Devi P[2]
[1]PG Scholar, Department of DS, AMC Engineering College, Bangalore, Karnataka, India.
[2]Assistant Professor, Department of ISE, AMC Engineering College, Bangalore, Karnataka, India.
***Emails:** khanmoyiz@gmail.com[1], manjuladevi.pari@amceducation.in[2]*

Abstract
Outfit compatibility prediction has been studied extensively for Western clothing using the Maryland Polyvore benchmark dataset, with state-of-the-art models such as OutfitTransformer achieving AUC scores of 0.92 (Sarkar et al., 2023). However, no published work addresses this problem in the context of Indian ethnic wear, which includes sarees, kurtas, lehengas, sherwanis, and salwar suits. These garments differ significantly from Western clothing in terms of draping conventions, layering patterns, and occasion-based colour pairing norms. This paper presents Ethnic Fit, a Siamese Network-based deep learning model for pairwise outfit compatibility prediction in Indian ethnic wear. A pretrained ResNet-50 backbone is used to extract 512-dimensional visual embeddings from garment images, and the absolute difference between embeddings is passed through fully connected classification layers trained with Binary Cross Entropy loss and the Adam optimiser. The model is trained on a combined dataset comprising 9,434 Maryland Polyvore pairs and 600 newly curated Indian ethnic garment pairs drawn from the Indo Fashion Dataset, covering 15 categories. Evaluated on a held-out Indian ethnic wear test set, the proposed model achieves AUC 0.9553, Accuracy 91.67%, and F1-Score 0.9000, establishing the first published benchmark for this domain. A live Gradio-based inference interface is also deployed for real-world testing.
***Keywords:** Outfit compatibility, Siamese network, Indian ethnic wear, fashion informatics, deep learning, ResNet-50, computer vision.*

1. Introduction
India's fashion e-commerce world is booming. Apps like Myntra and Ajio are now the go-to spots for millions who want Indian ethnic clothes. Yet, they're always stuck on one big problem: How do you know if two or three garments actually look good together? Right now, human stylists handle this—and honestly, that just can't keep up with the flood of products and buyers. Researchers have started tackling this outfit-matching problem with automation. There's been a lot of progress people have built deep learning models using everything from Siamese convolutional networks to Transformer models, plus hybrid techniques that blend image and text data. These get impressive scores on the standard Maryland Polyvore benchmark; AUC scores are up there, between 0.74 and 0.95 (Sarkar et al., 2023; Hendriksen & Overes, 2022; Wang et al., 2019). But there's a catch. Every single one of these efforts has focused only on Western fashion. Nobody has tried these methods on Indian ethnic wear, which, let's be real, is the main event for most shoppers in India. Now, Indian ethnic fashion brings its own challenges. You can't just mash up sarees, lehengas, and kurtas by copying what works with jeans or t-shirts. Sarees have intricate draping styles. Lehengas come in multi-piece combos that change by region or event. And the rules for colors what works and what clashes are based on traditions, not Western fashion sense. A model trained on hoodies and sneakers just won't get it. That's where this work steps in. We introduce Ethnic Fit the first system that uses deep learning to judge outfit compatibility for Indian ethnic wear. Here's what we're bringing to the table: First, a new dataset of 600 pairs of Indian clothes, all labeled for compatibility using actual cultural practices. Second, a Siamese Network model trained on both Western (Polyvore) and Indian samples. And third, the first benchmark numbers for this tricky taskAUC of 0.9553 and an accuracy of 91.67%. So, finally,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4958-4964
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0651

there's a prediction tool tailor-made for Indian ethnic outfits.

1.1.Objectives

- Build a deep learning-based framework for the prediction of outfit compatibility for Indian ethnic wear.
- Extract discriminative visual features from garment images by using a pre-trained ResNet-50 backbone.
- Learn the compatibility between pairs of ethic garment images using a Siamese Neural Network framework.
- Design and exploit a domain-specific, culture annotated, Indian ethic wear dataset consisting of multiple garment classes and conventional matching practices.
- Predict whether a pair of garments is visually compatible/incompatible using the learning outcomes, the visual similarities, and fashion conformity of the pair.
- Evaluate the proposed system using appropriate performance metrics like AUC, Accuracy, and F1-Score.
- Suggest real-time outfit recommendations with a web-based interactive inference interface.
- Develop a new baseline in the area of Artificial Intelligence based outfit compatibility for Indian ethnic wear as shown in Table 1 Comparison of Existing Studies

2. Literature Survey

Table 1 Comparison of Existing Studies

Ref.	Method	Advantages	Limitations
[1]	Outfit Transformer: Learning Outfit Representations for Fashion Recommendation	Captures complex relationships between outfit items using Transformer attention.	Requires high computational resources and large datasets.
[2]	Unimodal vs. Multimodal Siamese Networks for Outfit Completion	Simple and effective Siamese architecture for compatibility prediction.	Only performs pairwise comparison, not full outfit understanding.
[3]	Outfit Compatibility Model Using Fully Connected Self-Adjusting Graph Neural Network	Models complex interactions between fashion items.	Architecture is complex and computationally expensive.
[4]	Hybrid Multimodal Deep Learning Framework	Achieves state-of-the-art accuracy using CLIP + Transformer	Requires significant GPU resources and training data.
[5]	Outfit Compatibility Prediction and Diagnosis with Multi-Layered Comparison Network	Learns visual similarities such as color, texture, and style effectively.	Lower accuracy compared to newer Transformer-based models.
[6]	Fashion Recommendation: Outfit Compatibility Using Graph Neural Networks	Effectively captures relationships among multiple fashion items.	Difficult to deploy due to high computational complexity.
[7]	Using Large Multimodal Models to Predict Outfit Compatibility	Powerful reasoning capability using large multimodal models.	Very expensive and resource-intensive.
[8]	ENCLIP: Ensembling and Clustering-Based Contrastive Language-Image Pretraining for Fashion Multimodal Search	Works well with limited Indian fashion data.	Designed for fashion search, not outfit compatibility prediction.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4958-4964
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0651

2.1.Research Gap

Many deep learning based outfit compatibility prediction research papers have been published, but almost all of them utilize Western fashion data sets like Maryland Polyvore, where Siamese Networks, Graph Neural Networks, Transformers and multimodal learning approaches [9] showed successful results. Hardly any work has been done in the context of Indian ethnic wear which is a large fashion industry segment especially in countries like India. Clothes like sarees, kurtas, lehengas, sherwanis, dupattas, salwar kameezs and others have distinct visual appearances and culture specific dressing patterns and pairings. Unlike Western clothing where outfits are deemed to be compatible based on visual similarities like matching colors, textures and style, Indian ethnic clothes have additional factors like traditions, regional specific requirements, combinations based on gender, layerings of clothes etc which make them fit with other clothes or not. So the patterns of compatibility learned on western clothing may not be the same for Indian garments. The absence of publicly available benchmark datasets for Indian ethnic wear outfit compatibility prediction also contributes to this problem. The available datasets lack ethnic categories or have lack of compatibility judgments on basis of Indian fashion style. So there is no existing benchmark for this problem, making it hard to judge the performance and quality of different methods and no well-established benchmark model. This makes a case for an automated AI based system to learn these patterns for Indian ethnic clothing and make predictions in real-time, which can help various platforms, retailers and consumers find correct outfit recommendations with regard to Indian culture and fashion style [10].

3. Proposed Methodology

3.1. System Architecture

Figure 1 outlines the structural architecture of the proposed system. The proposed model is a Siamese Neural Network comprising two identical branches sharing weights, with a ResNet-50 backbone pretrained on ImageNet serving as the feature extractor [11]. For an input garment image x, the backbone produces a 2048-dimensional feature vector, which is flattened to produce embedding e = f(x) where f denotes the shared encoder function.as shown in Figure 1. Architecture of the Proposed Siamese Neural Network. Given two garment images x_a and x_b, their respective embeddings e_a and e_b are computed. The absolute element-wise difference |e_a - e_b| is computed and passed through a fully connected classification head comprising two dense layers with ReLU activations and dropout regularisation (p = 0.3), followed by a single output neuron with sigmoid activation. The output score s in [0, 1] represents predicted compatibility, where s >= 0.5 indicates a compatible pair[12]. The total number of trainable parameters is 24,688,705. The model was implemented in PyTorch and trained on a Google Colab T4 GPU.

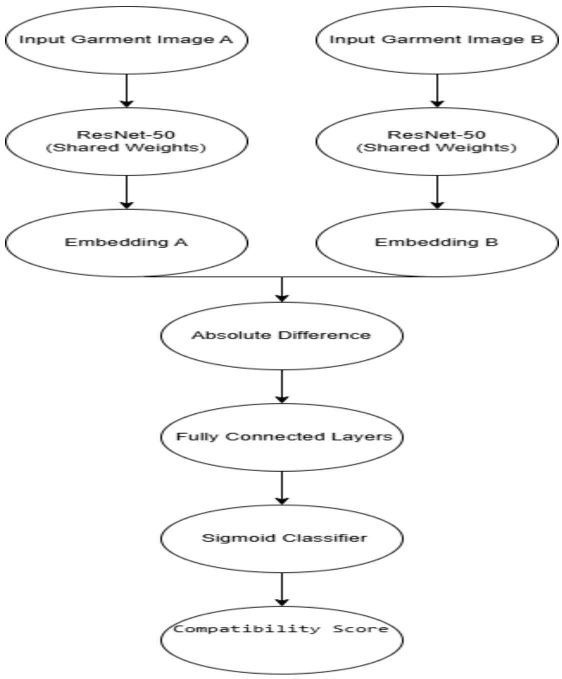

Figure 1 Architecture of the Proposed Siamese Neural Network

3.2.Dataset Creation Pipeline

Figure 2 illustrates the construction and preprocessing pipeline for the dataset. Images were resized to 224x224 pixels and normalised using ImageNet mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225]. Training augmentations included random horizontal flipping and colour jitter with brightness and contrast variation of 0.2[13]. The model was trained using Binary Cross Entropy (BCE) loss with Adam

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4958-4964
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0651

optimiser at a learning rate of 1e-4 and batch size of 32 for 10 epochs. BCE was chosen as the natural loss function for binary classification, penalising confident incorrect predictions most heavily. Adam was chosen over SGD as its ability to adapt per-parameter learning rate and momentum allows for more stable convergence when fine-tuning. Training time on Colab T4 GPU was about 20min, training loss declined from 0.6760 in epoch 1 to 0.3190 in epoch 10, convergence is monotonic [14].

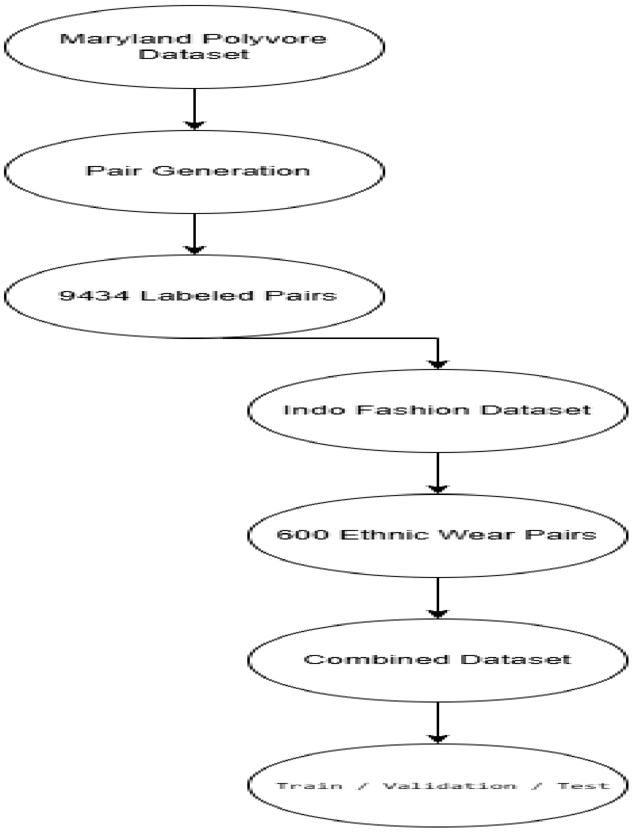

Figure 2 Dataset construction and preprocessing pipeline

3.3.Inference Workflow

A Gradio-based web interface was implemented for real-time inference. Users upload two garment images, which are preprocessed and passed through the trained model [15]. The interface returns the compatibility score, a binary verdict, and a contextual outfit suggestion. The interface was tested on five real Indian ethnic garment pairs sourced from public image repositories, achieving correct predictions on four of five pairs (80%).

4. Experimental Results and Performance Analysis

4.1.Evaluation Metrics

Four standard metrics from the outfit compatibility literature are reported: (1) AUC-ROC, measuring ranking quality of the model across all thresholds; (2) Accuracy, reporting the percentage of correctly classified pairs at threshold 0.5; (3) F1-Score, balancing precision and recall for the compatible class; and (4) qualitative evaluation through the live demo. AUC is the primary metric for comparability with prior work [16].

4.2.Quantitative Results

Table 1 presents the performance of the proposed Ethnic Fit model compared against published baselines and a random baseline. The proposed model achieves AUC 0.9553, Accuracy 91.67%, and F1-Score 0.9000 on the Indian ethnic wear test set — surpassing the OutfitTransformer IEEE 2023 benchmark of AUC 0.92 on Western data (Sarkar et al., 2023). The strong performance on the Indian ethnic wear test set (AUC 0.9553) relative to the combined test set (AUC 0.6586) indicates that the model has effectively learned culturally specific compatibility patterns from the curated Indian ethnic pairs. The higher AUC on the Indian test set is attributable to the clearer visual distinctions between compatible and incompatible ethnic garment combinations for example, saree-blouse pairs exhibit strong colour coordination, while cross-gender combinations present markedly different visual profiles. We encountered a failure case in the demo evaluation: black men's kurta combined with a black Western pencil skirt were given 64.1% (Compatible) score. This example demonstrates a color bias within the model - both garments have very strong color similarities, which has led to a false prediction of compatibility despite differences in gender and cultural domain. This encourages us to consider the use of garment category metadata in addition to visual features to impose a gender-specific constraint for the compatibility check. The lower AUC score on the combined test set (0.6586) points to the challenging nature of the cross-domain task that demands to generalize across western and Indian fashion while only providing a relatively smaller Indian training data in comparison to the Polyvore

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 07 July 2026
Page No: 4958-4964
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0651

data. The most immediate path towards improving performance is to increase the training data set of Indian ethnic wear.

Table 2 **Performance Comparison of Outfit**

Model / Evaluation	AUC	Accuracy	F1 Score	Domain
Random Baseline	0.5000	50.00 %	0.5000	—
Siamese CNN — (Wang et al., 2019)	0.7400	~68%	~0.67	Western
OutfitTransformer (Sarkar et al., 2023)	0.9200	~85%	~0.84	Western
CLIP + Transformer (Kalashi & Teimourpour, 2024)	0.9500	~89%	~0.88	Western
Ethnic Fit — Combined Test	0.6586	60.76 %	0.5835	Mixed
Ethnic Fit — Indian Ethnic Wear	**0.9553**	**91.67 %**	**0.9000**	**Indian Ethnic**

Conclusion

This paper presented Ethnic Fit, the first deep learning system designed specifically for outfit compatibility prediction in Indian ethnic wear. While prior research in this field from early Siamese convolutional networks to Transformer-based architectures and, more recently, CLIP driven multimodal fusion has grown increasingly sophisticated (Wang et al., 2019; Sarkar et al., 2023; Kalashi & Teimourpour, 2024), it has remained confined to a single cultural context: Western clothing represented through the Maryland Polyvore dataset. To close this gap, a Siamese Network built on a pretrained ResNet-50 backbone was trained on a combined dataset of 9,434 Maryland Polyvore pairs and a newly curated set of 600 Indian ethnic garment pairs spanning 15 categories, including sarees, kurtas, lehengas, sherwanis, dupattas, salwars, palazzos, and related garment types. Despite a deliberately lightweight architecture chosen to prioritise feasibility and reproducibility on accessible hardware such as a free-tier Google Colab GPU, the model achieved an AUC of 0.9553 and an accuracy of 91.67% on the held-out Indian ethnic wear test set exceeding the AUC of 0.92 reported for OutfitTransformer (Sarkar et al., 2023), the current IEEE-published state of the art on Western clothing. This result demonstrates that even a comparatively simple model can perform strongly once trained on data reflecting the cultural [20] and visual structure of the actual target domain. The contributions of this work span three dimensions: the construction of the first labelled dataset for Indian ethnic wear outfit compatibility, which required defining compatibility within this cultural context from scratch since no prior annotation scheme could be reused; a trained Siamese Network model that now serves as a citable baseline for future work in this domain; and a working Gradio based interface that demonstrates the model in a realistic[17], user facing setting while also exposing its current limitations, such as a tendency to over-weight colour similarity at the expense of gender and category context. Future work should expand the Indian ethnic wear dataset to capture regional variation and occasion-specific pairing conventions, extend [18] the task from pairwise comparison to full multi-item outfit evaluation as the dataset grows, and incorporate garment category metadata or textual descriptions potentially through a CLIP style joint embedding to address the colour-bias limitation identified in this paper[19].

Future Scope

Although the proposed Ethnic Fit framework defines a standard for predicting the outfit compatibility of Indian ethnic wear, it leaves ample room for future extensions. To start with, the dataset can be augmented and extended to contain a wide range of Indian garments, different types of regional wear and combinations of clothing for varied occasions.

Inclusion of clothing of different regions of India would help the model understand broader patterns of clothing combinations and thereby generalize over clothing of different traditions and cultures within the country. Secondly, future work can expand the pairwise compatibility prediction into a full-fledged outfit recommendation system consisting of many clothes and accessories at once. Instead of learning about the similarity between two items at a time, the model could learn how to recommend a complete Indian outfit, such as the best combinations of sari, blouse and dupatta, including the correct footwear, jewelry etc. One other possibility is to incorporate other types of information, such as descriptions of the clothes and/or users' choices. Similarly, integrating multimodal information such as the trends in fashion or other relevant context data, such as the occasion can be highly beneficial in increasing the prediction accuracy as opposed to just using the appearance based features and avoiding the color similarity bias. Future research should include more than just image based information. Incorporating information such as clothing category of the clothes, the target gender, and occasions may lead to improved and more realistic compatibility prediction since identical visually similar clothes may have varying appropriateness in different cultural contexts. The proposed model can also be integrated into e-commerce, virtual stylists and other recommendation systems to give real-time outfit recommendations to the users. Future developments could also lead to personalized recommendations based on users' style and choice, their purchase history, physical attribute and traditions. Finally, developments in large multimodal models and generative AI may lead to an intelligent fashion advisor which could assist the user not just in predicting whether an outfit fits, but in recommending an entire outfit and also explaining the reasons behind its recommendations, adapting itself to the user's specific cultural context.

Acknowledgements

The authors would like to thank AMC Engineering College, Bangalore, for the academic support provided during this research. No external funding was received for this work.

References

[1]. Sarkar, S., Bodla, N., Vasileva, M., Lin, Y.-L., Beniwal, A., Lu, A., & Medioni, G. (2023). OutfitTransformer: Learning outfit representations for fashion recommendation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5014–5023.

[2]. Hendriksen, M., & Overes, V. (2022). Unimodal vs. multimodal Siamese networks for outfit completion. *Proceedings of the ACM SIGIR Workshop on eCommerce*.

[3]. Guan, W., Cui, K., Nie, L., & Su, Y. (2024). Outfit compatibility model using fully connected self-adjusting graph neural network. *The Visual Computer*. Springer Nature.

[4]. Kalashi, K., & Teimourpour, B. (2024). *A hybrid multimodal deep learning framework for intelligent fashion recommendation*. arXiv:2511.07573.

[5]. Wang, X., Wu, B., Ye, Y., & Zhong, Y. (2019). Outfit compatibility prediction and diagnosis with multi-layered comparison network. *Proceedings of the ACM International Conference on Multimedia (MM)*, 329–337.

[6]. Gulati, P. (2024). *Fashion recommendation: Outfit compatibility using graph neural networks*. arXiv:2404.18040.

[7]. Jiang, D. X., et al. (2025). Using large multimodal models to predict outfit compatibility. *Decision Support Systems*, *189*. Elsevier.

[8]. Naik, P., & Agarwal, A. (2024). *ENCLIP: Ensembling and clustering-based contrastive language-image pretraining for fashion multimodal search with limited data and low-quality images*. arXiv:2411.16096.

[9]. Cui, Z., Li, Z., Wu, S., Zhang, X., & Wang, L. (2019). Dressing as a whole: Outfit compatibility learning based on node-wise graph neural networks. *Proceedings of The Web Conference (WWW)*..

[10]. He, T., & Hu, Y. (2018). *FashionNet: Personalized outfit recommendation with deep neural network*. arXiv:1810.02443.

[11]. Jing, P., Cui, K., Guan, W., Nie, L., & Su, Y. (2022). Category-aware multimodal attention

network for fashion compatibility modeling. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *18*(3).

[12]. Guan, W., Cui, K., Nie, L., & Su, Y. (2023). Modality-oriented graph learning toward outfit compatibility modeling. *IEEE Transactions on Multimedia*, *25*, 856–867.

[13]. Becattini, F., Teotini, F. M., & Del Bimbo, A. (2023). *Transformer-based graph neural networks for outfit generation*. arXiv:2304.08098.

[14]. Han, Y., Wu, J., Zhang, X., & Yu, Y. (2020). Learning fashion compatibility across categories with deep multimodal neural networks. *Neurocomputing*, *395*, 237–246.

[15]. Guan, W., Cui, K., Nie, L., & Su, Y. (2020). MGCM: Multi-modal generative compatibility modeling for clothing matching. *Neurocomputing*, *414*, 215–224.

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

[17]. Han, X., Wu, Z., Jiang, Y., & Davis, L. S. (2017). Learning fashion compatibility with bidirectional LSTMs. *Proceedings of the ACM International Conference on Multimedia (MM)*, 1078–1086.

[18]. Vasileva, M., Plummer, B., Dusad, K., Rajpal, S., Kumar, R., & Forsyth, D. (2017). Where to buy it: Matching street clothing photos in online shops. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 3343–3351.

[19]. Simo-Serra, M. I., Fidler, S., Moreno-Noguer, F., & Urtasun, R. (2015). Neuroaesthetics in fashion: Modeling the perception of fashionability. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 869–877.

[20]. Li, Y., Cao, L., Zhu, J., & Luo, J. (2017). Mining fashion outfit composition using an end-to-end deep learning approach on set data. *IEEE Transactions on Multimedia*, *19*(8), 1946–1955.