Spectrum of Engineering Sciences
ISSN (e) 3007-3138 **(p)** 3007-312X **Volume 4, Issue 5, 2026**

MACHINE LEARNING BASED PREDICTION OF DERMATOPHYTOSIS: A COMPARATIVE ANALYSIS OF THE DERMNET DATASET

Pirah Daniyal*[1], Dua Abro[2], Farah Leghari[3], Aneeqa Khanzada[4], Kashaf Khanzada[5]

*[1,2,4,5]*Liaquat University of Medical Health and Sciences,*
[3]*Fast Nuces, Karachi*

*[1]lagharipirah3@gmail.com, [2]dua.abro2001@gmail.com, [3]farahlaghari222@gmail.com, [4]aneeqakhanzada1@gmail.com, [5]khanzadakashaf@gmail.com

DOI:https://doi.org/10.5281/zenodo.20507653

Keywords

Article History
Received: 02 April 2026
Accepted: 12 May 2026
Published: 30 May 2026

Copyright @Author
Corresponding Author: *
Pirah Daniyal

Abstract

Dermatophytosis, commonly known as ringworm or tinea, is one of the most spread fungal skin infections all over the world, affecting nearly 20–25% of the population on earth. When it comes to successful therapies, it is difficult to diagnose the problem promptly and correctly. Although it is still difficult to explain the difference among fungal and non-fungal skin disorders because they look so almost similar. This study introduces a computer-based recognition system for dermatophytosis by using a hybrid methodology that combines old feature extraction techniques with deep learning-based transfer learning. We used DermNet dataset downloaded from Kaggle which is free for everyone to create an equal amount of differentiation tasks that differentiate fungal infections from 21 non-fungal dermatological classes. For bringing an image to a standard resolution of 224×224 and bringing them down to the [0,1] range of values, preprocessing was done. There are multiple methods available for obtaining features, but in this study these two are used: (1) an old procedure that makes a combination of Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) descriptors, and (2) for deep features, a ResNet50 network was used. We used 4 machine learning classification algorithms on the deeper attributes and also on a manually created one: Support Vector Machine (SVM) with RBF kernel, Random Forest, XGBoost, and K-Nearest Neighbours (k=5). A KNN algorithm trained sequentially by using a modified ResNet50 CNN and a custom classification head. Accuracy, precision, recall, F1-score, and ROC-AUC are used to test the simulations on an assigned set of tests (15% split). Our results are showing the efficiency of the combined methodology, with the modified ResNet50 CNN having higher performance levels. This work gives an applicable and scalable structure to facilitate the automation of fungus dermatological infection assessment.

INTRODUCTION:

Dermatophytosis are fungal infections that affect the epidermis, strands of hair, and nails, which are all hydrated tissues [1, 10, 11, 13, 18] Dermatophytes from the species Microsporum, Trichophyton, and Epidermophyton, are mostly responsible for these infections [10, 11]. Dermatophytosis is very common in tropical and subtropical areas because of the heat, moisture levels, inadequate sanitation, too many people, and extended periods of excessive sweating [2]. Asexual infections caused by fungi have been documented globally, impacting approximately twenty to twenty-five percent of the population worldwide [3, 19, 22 ,23]. Tinea corporis, tinea cruris, tinea pedis, tinea capitis, and onychomycosis are all prevalent kinds of dermatophytosis [4, 20].

The clinical manifestation of the condition has undergone substantial changes in the past few years, especially in South Asia. New types of fungi, like Trichophyton indotineae, are becoming more widespread in countries like India, Pakistan, Bangladesh, and Nepal [5]. The fungus Trichophyton indotineae is particularly alarming due to a correlation with persistent infections, significant cutaneous engagement, and resistant status to frequently utilized antifungal agents like terbinafine [6, 21]. The increase in the species of fungus causing ringworms is resistant creating hurdles for the healthcare providers treating patients using old methods of treatments as a result infections and its spread is becoming common among communities [7].

Fungal diseases are becoming more vulnerable due to the weather changes and extremely hot temperature which are the effects of global warming [8]. Fungi can more easily spread when the temperature is high, moisture is high, uneven patterns of rain, larger portion of hot weather [9]. Dermatophytes flourish in humid and hot conditions, rendering areas with high populations of South Asian countries particularly susceptible to a heightened occurrence of dermatophytosis [12]. Environment-related variables like humidity, precipitation, temperature, urban development, and excessive population growth may be significant indicators of the risk of infections caused by fungi [2].

Artificial intelligence has become an indispensable tool in the medical field due to the fact it can look at a lot of different factors at once and find underlying trends in complicated datasets [13, 15, 16, 17]. Supervised approaches to learning are especially helpful because they are able to determine how patient characteristics relate to disease results derived from information that has been categorized [14, 16, 17]. For predicting dermatophytosis, machine learning algorithms can use socioeconomic, medical, and environmental parameters like age, sex, moisture, temperature, perspiration, dermatitis, acne, normal scalability, and history of relatives to guess how likely it is that someone will get the disease [17, 24, 25].

Even though the condition known as dermatophytosis is progressively more common in South Asia, there are still not many large long-term datasets that include individual health care data as well as data on the surroundings and the weather. The majority of research that are accessible only look at small groups of people in hospitals, which makes it hard to apply the results to other areas. So, we need to make machine learning algorithms that operate well regardless of whether there isn't much data available.

This study seeks to introduce a supervised learning methodology for forecasting dermatophytosis resulting from warming temperatures, utilizing limited healthcare and ecological datasets. The research concentrates on determining relevant databases, preliminary processing techniques, feature identification methodologies, machine learning algorithms, and testing metrics for forecasting fungal infections, specifically Trichophyton indotineae, within South Asian populations.

Methodology

In the section of methodology sequence of work is given from the selection of dataset till the results which were obtained from the trained model.

1. Selection of Dataset and Preparation

The DermNet dataset obtains a good collection of dermatological images consisting of 23 distinct disease classes. Images that are used are taken from DermNet NZ, which is a freely available dermatological resource preserved by the New Zealand Dermatological Society.

Two different classes that were chosen as favourable (Fungal) were (1) Tinea Ringworm Candidiasis and other Fungal Infections and (2) Normal Nail Microbes and additional Nail Infections. The remaining 21 courses were all assigned as negative (non-fungal). random under-sampling strategy was used on the non-fungal pool to cope up with the imbalances of class. This means if a number is picked from not-fungal pictures that was the same as the total number of fungi-based images. random seed 42 was employed, to get the surety of not fungal pictures. The data was split the appropriate dataset into three parts: 70% for training, 15% for confirmation, and 15% for testing as shown in table 1. The distribution makes it clear that all three divisions had the same number of classes. for loading files same way in different experiments, we saved splits as NumPy .npy files.

Table 1: Dataset split summary (approximate values based on dataset size)

Split	Images	Fungal	Non-Fungal	Proportion
Training	~2,100	~1,050	~1,050	70%
Validation	~450	~225	~225	15%
Test	~450	~225	~225	15%
Total	~3,000	~1,500	~1,500	100%

2. Image Preprocessing

Each image that is present was processed through a standard four steps using OpenCV: Read: Images were loaded from the disk and successfully skipped all the unwanted images. Resize: ResNet50 and other compatible systems were used for standard input resolution before that images were resized to 224 x 224. Conversion of Colour: The BGR channel order in OpenCV was changed to RGB to fit with how deep learning is usually done.

This preliminary processing workflow makes sure that graphics with different initial dimensions and file types (JPEG, PNG) are all the same. It also gets them ready for both manually extracted features and input directly into neural network models.

3. Feature Extraction

Feature extraction was done by both traditional and modern methods. Histogram of Oriented Gradients (HOG) was used to detect the shape and edge patterns available in the epidermis lesion photographs. First of all, every RGB image was changed into grayscale. Then, the features of HOG were taken out using 9 gradient orientation bins, 16 × 16-pixels, and 2 × 2 cell blocks. This whole procedure gives a detailed vector feature for all images, which helps model differentiate common fungal lesion patterns like as round shapes, scaling, and uneven borders.

Local Binary pattern was used to explain the texture of the epidermis surface. It works on comparison logic each pixel with pixels next to it. In this study, 8-bit binary code was generated by using 8-neighbouring pixels with a radius of 1. finally, the histogram created for feature vector. LBP was proven useful for the identification of unsmooth surface, which is common symptom of infections caused by dermatophyte.

One complete image was created by using the combination of the HOG and LBP feature vectors. This combination of feature vector consists of both edge and shape information coming from HOG, as well as information of texture from LBP. use of combination of method gives better performance also gives good results.

ResNet50 model was used for deep feature extraction. Without final classification layer model was loaded and it used global average pooling for the generation of 2048 dimensional feature vector. While this stage is going on, all other layers of the network remain frozen. Pré processing done by using ResNet50 before extraction of feature. later wards these features were used as input to KNN for more specific patterns visually for large dataset.

Classification Models

Following are the five classification models which were used for the training of the model.

1. Support Vector Machine (SVM)

A Support Vector Machine model with an RBF (Radial Basis Function) kernel was trained using the combined HOG and LBP feature vectors. Before training, the data was standardized using StandardScaler so that all features had similar importance. The model was trained with C = 1.0, gamma = "scale", and probability = True so that probability scores could later be used for ROC curve analysis. Training the SVM model on the full dataset took around 2 to 3 minutes.

2. Random Forest

The Random Forest model was trained using 200 decision trees on the combined HOG and LBP features. Unlike SVM, this model did not require feature scaling. Random Forest works by combining the predictions of many decision trees, which helps improve accuracy and reduce the effect of noise in the dataset. For making faster training Parallel processing was used with n_jobs = -1.

3. XGBoost

200 estimators were used to train the XGBoost model, a highest tree depth of 6, and a rate of learning was 0.1. Also, subsample and colsample_bytree values of 0.8 were utilized for the reduction of overfitting. The set of data for validation given during training for making model to observe log-loss also make its performance better with time. This model works by correcting mistakes slowly that has been made in previous tree which is its strongest attribute as machine learning model.

4. K-Nearest Neighbors (KNN)

Main reason for using this model was extraction of features at depth by using ResNet50 model. Value of K was taken five also Euclidean distance was used. It already sure that before making model trained all the features are chosen as per standard. This model becomes more effective when ResNet50 is being utilised as it classifies same disease patterns together.

5. Fine-Tuned ResNet50 CNN

Fine tuning was done by the use of ResNet50 model for the classification of fungal disease. Additional layers were also added after the output for normalization, density and dropping something. Sequence that was used: Batch Normalization → Dense (256, ReLU) → Dropout (0.5) → Dense (128, ReLU) → Dropout (0.3) → Dense (1, Sigmoid).

On purpose 140 layers of ResNet50 model were kept frozen for the reduction training time and prevention of overfitting. Rest of the layers were fine tuned. Model training was done by Adam optimizer with learning rate 0.0001 with the loss of binary cross entropy. Size of batch was 32, 10 iterations were run for training. Stopping early was also applied to model so that it can stop once validation performance stops updating.

Results

For the Assessment of each model Accuracy, precision, recall, F1 score and ROC-AUC (area underneath the receiving operator character curve) were used. For the calculation ROC-AUC, statistical outputs (predict_proba) were used.

Figure 1a shows that the SVM model correctly classified 218 Non-Fungal and 277 Fungal cases. It performed better in identifying fungal infections than non-fungal cases.

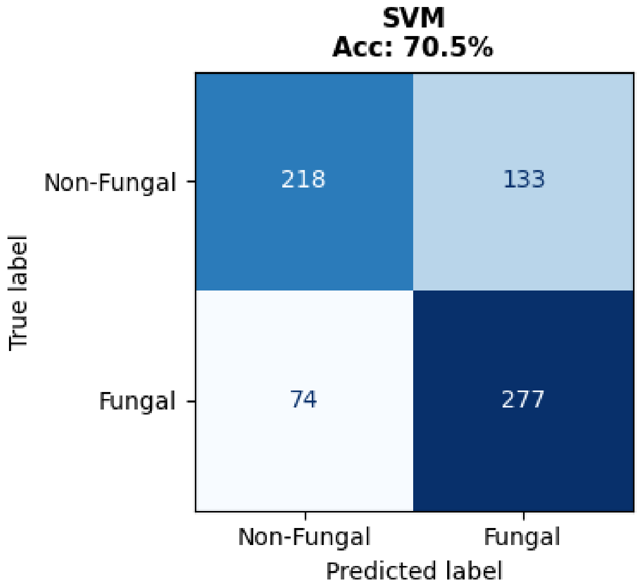


Figure 1(a): SVM (Accuracy: 70.5%)

Figure 1b shows Random Forest correctly identified 226 Non-Fungal and 248 Fungal cases. It had the highest number of false negatives, meaning it missed more fungal cases than other models.

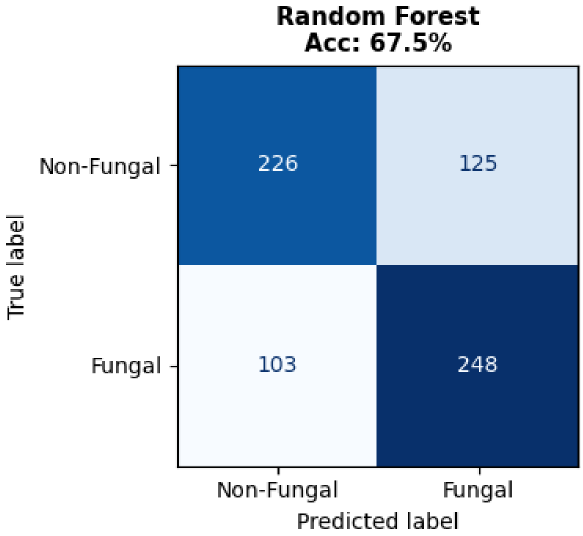


Figure 1(b): Random Forest (Accuracy: 67.5%)

Figure 1c shows XGBoost correctly classified 222 Non-Fungal and 259 Fungal cases. It performed slightly better than Random Forest and showed balanced results.

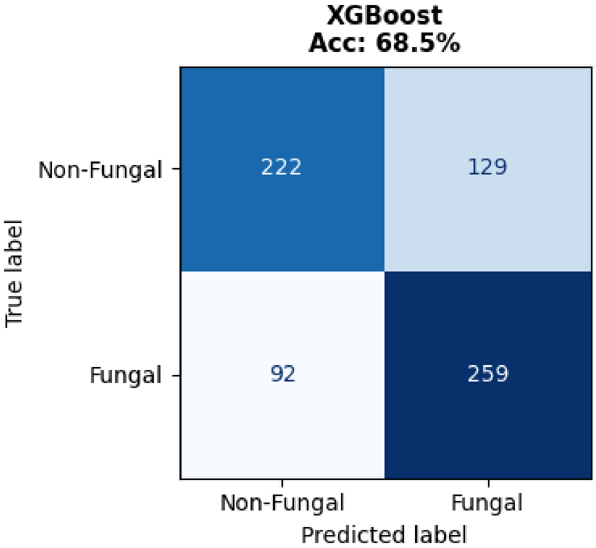


Figure 1 (c): XGBoost (Accuracy: 68.5%)

Figure 1d shows KNN achieved the highest accuracy. It correctly classified 290 Non-Fungal and 254 Fungal cases and had the lowest number of false positives.

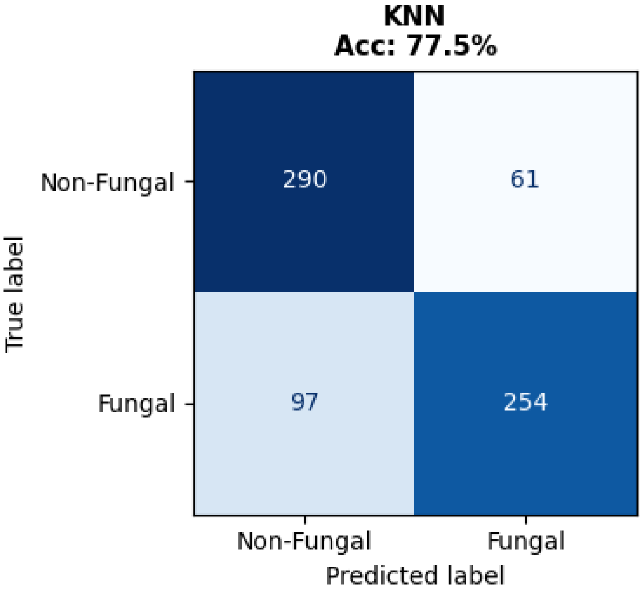


Figure 2 (d): KNN (Accuracy: 77.5%)

Figure 1e shows the fine-tuned ResNet50 CNN correctly classified 208 non-fungal and 270 Fungal cases. Its performance was lower than KNN, possibly due to the limited dataset size.

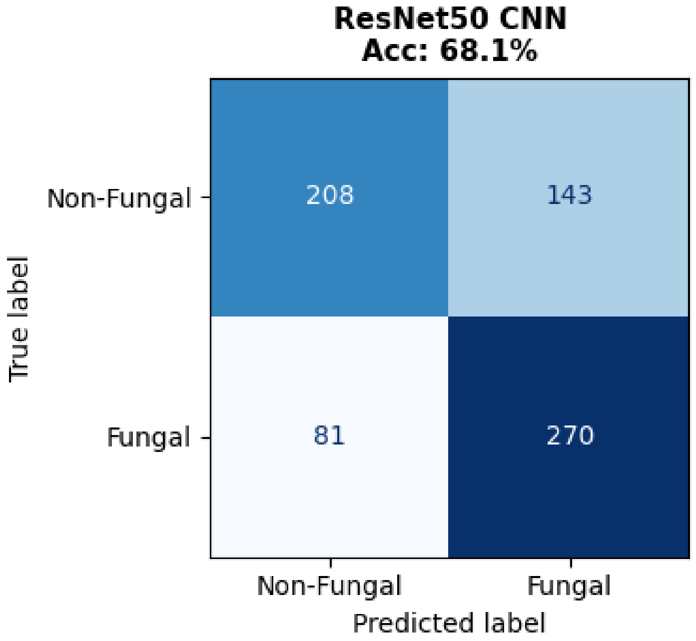


Figure 3 (e): ResNet50 CNN (Accuracy: 68.1%)

Figure 2 explains ROC curves for all models that how better they are identifying between the classes line that is farthest giving better results. KNN is farthest one at an AUC of 0.855 which that is 85.5% better classifying the model. While other models are performing within the range of 0.737-0.751, ever model is above the threshold.

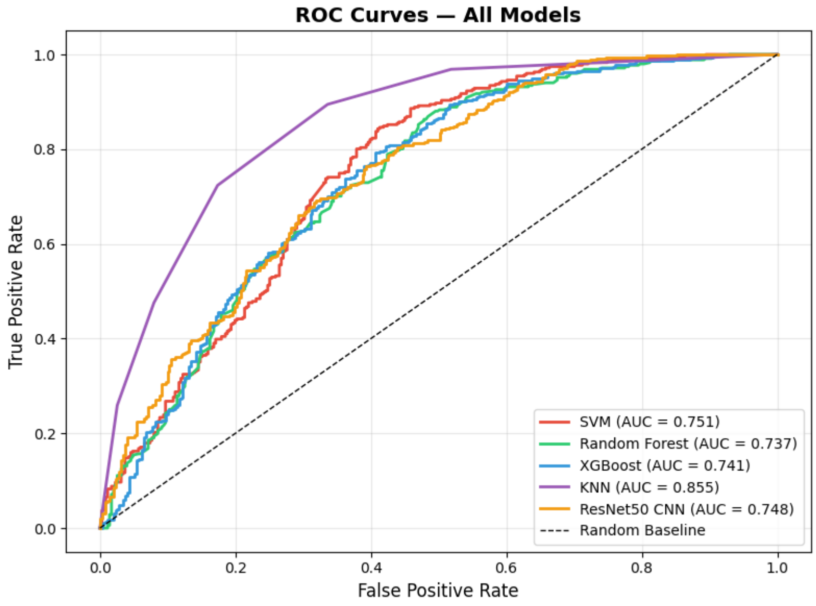
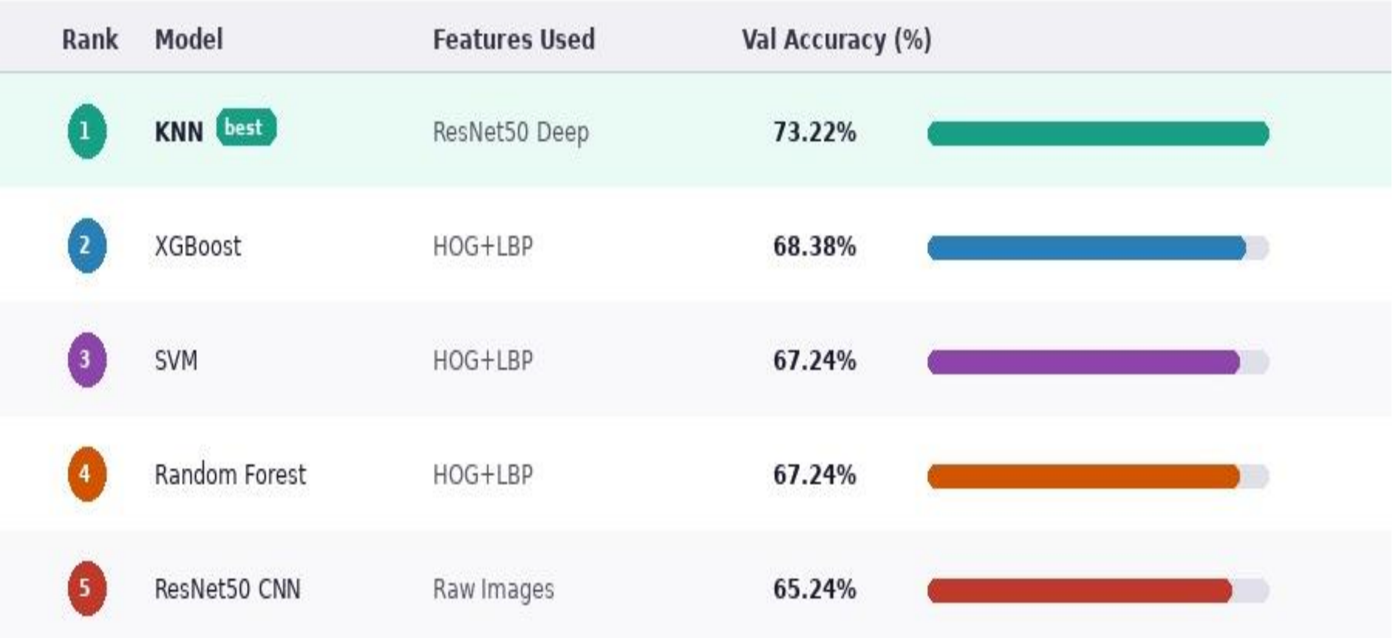


Figure 2: ROC curves of all models

Model Performance Comparison

Validation accuracy comparison of five classifiers trained on the DermNet dataset for dermatophytosis detection is shown in table 2 and in figure 3. KNN using ResNet50 deep features achieves the highest accuracy (73.22%), outperforming traditional HOG+LBP models (SVM, Random Forest, XGBoost) and the fine-tuned ResNet50 CNN (65.24%).

Table 2: Comparison of validation accuracy



Rank	Model	Features Used	Val Accuracy (%)
1	**KNN** best	ResNet50 Deep	**73.22%**
2	XGBoost	HOG+LBP	**68.38%**
3	SVM	HOG+LBP	**67.24%**
4	Random Forest	HOG+LBP	**67.24%**
5	ResNet50 CNN	Raw Images	**65.24%**

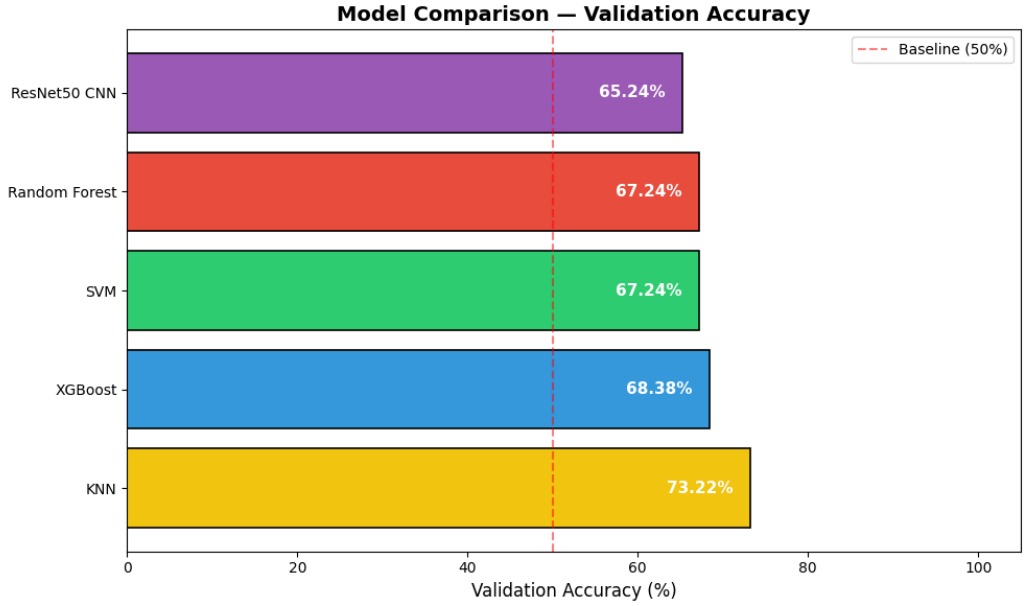


Figure 3: Graph of validation accuracy

KNN was best performing model among all in terms of learning with the accuracy of 77.49% and ROC-AUC of 85.48% on the other side traditional model like SVM, and XGboast were in a range of 73 to 79% accuracy. ResNet50 CNN was having accuracy of 65.38. Random forest was at the lowest as shown in table 3 and figure 4.

Table 3: Performance summary of test set

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Rank
KNN	77.49	80.63	72.36	76.28	85.48	1
SVM	70.51	67.56	78.92	72.80	75.08	2
ResNet50 CNN	68.09	65.38	76.92	70.68	74.75	3
XGBoast	68.52	66.75	73.79	70.09	74.10	4
Random Forest	67.52	66.49	70.66	68.51	73.70	5

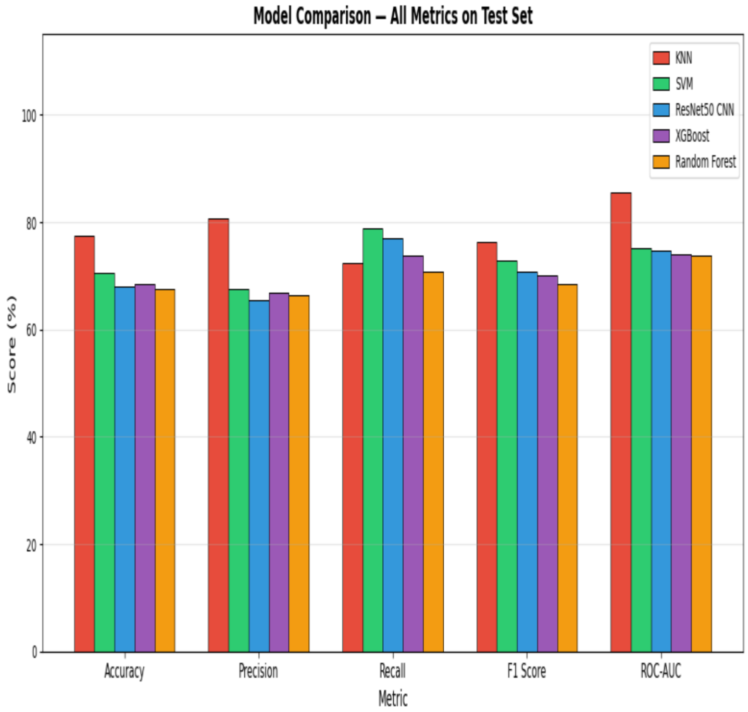


Figure 4: Model comparisons

Conclusion

This study utilized the DermNet dataset to demonstrate the capability of machine learning and deep learning models in distinguishing fungal skin infections from non-fungal dermatological conditions. The comparative analysis revealed that the ResNet50-based deep learning model achieved the highest overall performance, while K-Nearest Neighbors (KNN) using ResNet50-extracted deep features and XGBoost also produced promising results. The findings indicate that artificial intelligence-based approaches can effectively support the automated detection of dermatophytosis with high accuracy and reliability. Such systems have the potential to facilitate early diagnosis and timely treatment of fungal skin infections, regardless of their underlying causes, including the increasing prevalence associated with climate change and global warming.

REFERENCES

[1] Chang, Che-Cheng, et al. "Prevalence and risk factors of zoonotic dermatophyte infection in pet rabbits in northern Taiwan." Journal of Fungi 8.6 (2022): 627.

[2] Sharma, Ananya, and Somesh Gupta. "Skin disease in tropics: impacted by heat, humidity, and healthcare neglect." *Critical Thinking in Contemporary Dermatology: Cognitive Essays*. Singapore: Springer Nature Singapore, 2024. 229-250.

[3] Havlickova, Blanka, Viktor A. Czaika, and Markus Friedrich. "Epidemiological trends in skin mycoses worldwide." *Mycoses* 51 (2008): 2-15.

[4] Barac, Aleksandra, et al. "Dermatophytes: update on clinical epidemiology and treatment." *Mycopathologia* 189.6 (2024): 101.

[5] Chowdhary, Anuradha, et al. "The emergence and worldwide spread of the species Trichophyton indotineae causing difficult-to-treat dermatophytosis: a new challenge in the management of dermatophytosis." *PLoS pathogens* 18.9 (2022): e1010795.

[6] Gupta, Aditya K., et al. "Trichophyton indotineae: epidemiology, antifungal resistance and antifungal stewardship strategies." *Journal of the European Academy of Dermatology and Venereology* 40.1 (2026): 29-45.

[7] Langfeldt, Antonia, Jeremy AW Gold, and Tom Chiller. "Emerging fungal infections: from the fields to the clinic, resistant Aspergillus fumigatus and dermatophyte species: a one health perspective on an urgent public health problem." *Current Clinical Microbiology Reports* 9.4 (2022): 46-51.

[8] Reyes-Montes, María del Rocío, et al. "Impact of climate change on dermatophytosis." *The Impact of Climate Change on Fungal Diseases*. Cham: Springer International Publishing, 2022. 73-85.

[9] Courtellemont, Laura, et al. "Epidemiology of Trichophyton verrucosum infection in Rennes University Hospital, France: a 12-year retrospective study." *Medical mycology* 55.7 (2017): 720-724.

[10] Surendran, K. A. K., et al. "A clinical and mycological study of dermatophytic infections." Indian journal of dermatology 59.3 (2014): 262-267.

[11] Deng, Ruixin, Xiaowen Wang, and Ruoyu Li. "Dermatophyte infection: from fungal pathogenicity to host immune responses." Frontiers in Immunology 14 (2023): 1285887.

[12] Ramos Irizarry, Paola, Daniel FQ Smith, and Asiya Gusa. "Climate change impacts on environmental fungi: Human health and fungal disease." *One Health and Mycology* (2025): 67-101.

[13] Fidan, Faruk, et al. "Artificial Intelligence-Based Analysis of Microscopic Images of Dermatophyte Fungi." *2025 33rd Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2025.

[14] F. Grignaffini, F. Barbuto, L. Piazzo, M. Troiano, P. Simeoni, F. Mangini, G. Pellacani, C. Cantisani, and F. Frezza, “Machine learning approaches for skin cancer classification from dermoscopic images: A systematic review,” *Algorithms*, vol. 15, no. 11, p. 438, Nov. 2022.

[15] Noronha, Stephanie S., et al. "Deep learning-based dermatological condition detection: A systematic review with recent methods, datasets, challenges, and future directions." *IEEE Access* 11 (2023): 140348-140381.

[16] BalaKrisna, G., Abdul Hannan Abdul Mannan Shaikh, and Mohit Tiwari. "Artificial intelligence and nanotechnology in biosensors." *Handbook of research on advanced functional materials for orthopedic applications*. IGI Global Scientific Publishing, 2023. 47-64.

[17] Vayadande, Kuldeep, et al. "Innovative approaches for skin disease identification in machine learning: A comprehensive study." *Oral Oncology Reports* 10 (2024): 100365.

[18] Brescini, L.; Fioriti, S.; Morroni, G.; Barchiesi, F. Antifungal Combinations in Dermatophytes. *J. Fungi* **2021**, 7, 727. https://doi.org/10.3390/jof709072

[19] Zhan P, Liu W. The Changing Face of Dermatophytic Infections Worldwide. Mycopathologia. 2017 Feb;182(1-2):77-86. doi: 10.1007/s11046-016-0082-8. Epub 2016 Oct 25. PMID: 27783316.

[20] Moriarty B, Hay R, Morris-Jones R. The diagnosis and management of tinea. BMJ. 2012 Jul 10;345: e4380. doi: 10.1136/bmj. e4380. PMID: 22782730.

[21] Ebert, A.; Monod, M.; Salamin, K.; Burmester, A.; Uhrlaß, S.; Wiegand, C.; Hipler, U.-C.; Krüger, C.; Koch, D.; Wittig, F.; et al. Alarming India-wide phenomenon of antifungal resistance in dermatophytes: A multicentre study. Mycoses 2020, 63, 717–728.

[22] Knackstedt R, Knackstedt T, Gatherwright J. The role of topical probiotics in skin conditions: A systematic review of animal and human studies and implications for future therapies. Exp Dermatol. 2020 Jan;29(1):15-21. doi: 10.1111/exd.14032. Epub 2019 Sep 18. PMID: 31494971.

[23] Nigat, Tsedenya Debebe, Sitote, Tilahun Melak, Gedefaw, Berihun Molla, Fungal Skin Disease Classification Using the Convolutional Neural Network, Journal of Healthcare Engineering, 2023, 6370416, 9 pages, 2023. https://doi.org/10.1155/2023/6370416

[24] Hussain, S. I., & Toscano, E. (2024). An extensive investigation into the use of machine learning tools and deep neural networks for the recognition of skin cancer: Challenges, future directions, and a comprehensive review. Symmetry, 16(3), 366.

[25] Sun, J., Yao, K., Huang, G., Zhang, C., Leach, M., Huang, K., & Yang, X. (2023). Machine learning methods in skin disease recognition: a systematic review. Processes, 11(4), 1003.