

Hybrid Deep Learning Model for Skin Cancer Detection Based on CNN and LSTM

Rashmi Sinha¹, Anil Kumar Sagar¹

¹ Sharda University, Greater Noida, India
sinharashmi380@gmail.com, aksagar22@gmail.com

Abstract. Skin cancer is among the most frequently diagnosed malignancies worldwide, and its outcome depends strongly on how early melanoma and related lesions are recognised. Convolutional neural networks (CNNs) have reached specialist-level accuracy on dermoscopic images, yet they treat each image as a single static input and say little about how local lesion structures relate to one another. Hybrid models that couple a CNN feature extractor with a long short-term memory (LSTM) network have therefore attracted growing interest, because the recurrent stage can model dependencies across image patches, feature-map regions, multiple scales or successive examinations of the same lesion. This survey reviews CNN–LSTM hybrid architectures for skin cancer detection and classification published between 2017 and 2026. It summarises the clinical background and public dermoscopy benchmarks, explains the computational principles of the CNN and LSTM components, and proposes a taxonomy of hybrid designs according to how the input sequence is constructed and how the recurrent unit is configured (unidirectional, bidirectional or attention-augmented). Representative studies are compared in terms of backbone, recurrent unit, dataset and reported performance, and the comparison is used to identify recurring weaknesses: class imbalance, limited diversity of skin tones, weak justification of the “sequence” in single-image settings, inconsistent reporting and limited interpretability. The review concludes with research directions, including longitudinal lesion monitoring, transformer–recurrent hybrids, multimodal fusion with patient metadata, federated training and lightweight deployment at the point of care.

Keywords: Skin cancer · Melanoma · Dermoscopy · Convolutional neural network · Long short-term memory · Hybrid deep learning · Survey

1 Introduction

Skin cancer comprises melanoma and the far more common keratinocyte (non-melanoma) carcinomas, chiefly basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (SCC). Melanoma accounts for a minority of cases but for most skin-cancer deaths, and global cancer statistics continue to place it among the cancers whose incidence is rising in fair-skinned populations [39]. Survival is high when a melanoma is excised at an early, thin stage and falls sharply once it has spread, so the practical problem is one of timely and accurate recognition.

Dermoscopy, a non-invasive technique that magnifies the skin and reduces surface reflection, improves diagnostic accuracy over naked-eye examination, but its interpretation depends on training and experience. Automated analysis of dermoscopic and clinical images has therefore been studied for more than two decades. The field changed decisively when Esteva et al. [12] showed that a single CNN trained on about 130,000 clinical images could match board-certified dermatologists on keratinocyte carcinoma and melanoma classification, and when Haenssle et al. [15] reported that a CNN outperformed most of 58 dermatologists in a dermoscopic reader study. Systematic reviews have since documented a rapid growth of CNN-based skin lesion classifiers [2, 5], mirroring the wider adoption of deep learning for early disease detection across medical domains [4].

Despite these successes, a conventional CNN maps an image to a label in one feed-forward pass. It captures local texture and shape well, but long-range relations between distant parts of a lesion, such as asymmetry between halves, irregular pigment network at the border, or regression structures appearing in only one region, are learned only implicitly through progressively larger receptive fields. Recurrent neural networks, and in particular the LSTM [17], were designed to model dependencies within ordered sequences. When the output of a CNN is reorganised into a sequence of patch descriptors, feature-map rows or multi-scale views, an LSTM can integrate evidence across that sequence before a decision is taken. The same principle extends naturally to genuinely temporal data, such as repeated dermoscopic images of a lesion acquired during follow-up. Hybrid CNN–LSTM designs, first popularised for video and image captioning [10], have accordingly been adapted to dermatology in several recent studies [1, 18, 28, 31, 33, 45].

Existing reviews of skin lesion analysis concentrate on CNNs, transfer learning and, more recently, vision transformers [2, 5]; a focused treatment of recurrent hybrids is still scarce [21]. This paper addresses that gap. Its contributions are as follows:

- It summarises the clinical context and the public dermoscopy datasets on which CNN–LSTM models are trained and evaluated.
- It explains the CNN and LSTM building blocks and the ways in which their outputs are coupled.
- It proposes a taxonomy of CNN–LSTM hybrids based on sequence construction and recurrent configuration.
- It compares representative studies and critically examines the reported evidence.
- It identifies open challenges and outlines directions for future research.

The remainder of the paper is organised as follows. Section 2 describes the review protocol. Section 3 presents the clinical and technical background, and Section 4 reviews datasets and evaluation metrics. Section 5 introduces the taxonomy of hybrid architectures, Section 6 compares published studies,

Section 7 discusses open challenges and Section 8 outlines future directions. Section 9 concludes the paper.

2 Review Methodology

The survey follows a structured literature review protocol guided by four research questions (RQs):

- **RQ1:** Which datasets and evaluation protocols are used to develop CNN–LSTM models for skin cancer detection?
- **RQ2:** How do existing studies construct the sequence presented to the recurrent stage, and which recurrent configurations are used?
- **RQ3:** What performance is reported, and how comparable and reliable are the reported results?
- **RQ4:** Which limitations remain unresolved, and which directions are most promising?

Literature was retrieved from Scopus, IEEE Xplore, SpringerLink, ScienceDirect, PubMed and Google Scholar using combinations of the terms *skin cancer*, *skin lesion*, *melanoma*, *dermoscopy*, *CNN*, *LSTM*, *BiLSTM*, *recurrent* and *hybrid deep learning*. Peer-reviewed journal articles, conference papers and book chapters published in English between 2017 and 2026 were considered. A study was included when it (i) addressed detection, segmentation or classification of skin lesions from images and (ii) either combined convolutional and recurrent components or served as a widely cited CNN baseline against which hybrids are compared. Studies without an image-based model, duplicate versions of the same work and non-peer-reviewed preprints were excluded, except for dataset descriptor papers that are standard references in the field. The selected works were then coded by sequence-construction strategy, recurrent unit, backbone, dataset, number of classes and reported metrics.

3 Background

3.1 Clinical Context

Pigmented skin lesions are commonly grouped into melanocytic lesions (benign nevi and melanoma) and non-melanocytic lesions (BCC, actinic keratosis and intraepithelial carcinoma, benign keratosis, dermatofibroma and vascular lesions). Clinicians apply structured heuristics such as the ABCD rule (asymmetry, border, colour, dermoscopic structures) and the seven-point checklist, which score features distributed across the lesion. These heuristics are inherently relational: asymmetry compares halves of a lesion, border irregularity is assessed along its perimeter, and the significance of a colour depends on where it appears. This relational character is one of the main motivations for adding a sequence model on top of local CNN features.

3.2 Convolutional Neural Networks

A CNN applies learned filters to an input image $\mathbf{X}$ to produce feature maps

$$\mathbf{F}^{(l)} = \sigma(\mathbf{W}^{(l)} * \mathbf{F}^{(l-1)} + \mathbf{b}^{(l)}), \quad (1)$$

where $*$ denotes convolution, σ a non-linear activation and $\mathbf{F}^{(0)} = \mathbf{X}$. Pooling layers reduce spatial resolution and increase the receptive field. Architectures such as VGG [36], ResNet [16], Inception and EfficientNet [40] are the backbones most frequently used for skin lesions, almost always initialised with ImageNet weights and fine-tuned. The effectiveness of such transfer learning for image-based disease recognition has been demonstrated in domains as different as plant pathology [14,37], ophthalmology [3] and mammography [41]. Very deep residual networks were among the first to reach strong results on dermoscopic melanoma recognition [44], and region-based detectors extend the same feature extractors to lesion localisation [43]. Image enhancement networks, including transformer-assisted designs [38], are increasingly used as a preprocessing stage before classification.

3.3 Long Short-Term Memory Networks

An LSTM processes a sequence $\{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ by maintaining a cell state $\mathbf{c}_t$ regulated by input, forget and output gates [17]:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (2)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (3)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o), \quad (4)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (5)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t, \quad (6)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t), \quad (7)$$

where $\odot$ denotes element-wise multiplication. The additive update of $\mathbf{c}_t$ mitigates the vanishing-gradient problem of simple recurrent networks. A bidirectional LSTM (BiLSTM) [13] runs two LSTMs in opposite directions and concatenates their hidden states, so that each position is informed by both preceding and following context; BiLSTMs are well established in sequence tasks such as text classification [25]. The gated recurrent unit (GRU) [6] is a lighter alternative with two gates. LSTMs have also been benchmarked against classical learners such as SVM, naive Bayes and random forest in general predictive analytics [26].

3.4 Coupling CNN and LSTM

In a CNN–LSTM hybrid, the CNN acts as an encoder g_θ that maps each element of an input sequence to a descriptor, $\mathbf{x}_t = g_\theta(\mathbf{S}_t)$, and the LSTM aggregates the

descriptors. The final hidden state, or an attention-weighted combination of all hidden states,

$$\alpha_t = \frac{\exp(\mathbf{v}^\top \tanh(\mathbf{W}_a \mathbf{h}_t))}{\sum_{k=1}^T \exp(\mathbf{v}^\top \tanh(\mathbf{W}_a \mathbf{h}_k))}, \quad \mathbf{z} = \sum_{t=1}^T \alpha_t \mathbf{h}_t, \quad (8)$$

is passed to a softmax classifier $\hat{\mathbf{y}} = \text{softmax}(\mathbf{W}_y \mathbf{z} + \mathbf{b}_y)$. The general pipeline is shown in Fig. 1. The key design decision is what $\mathbf{S}_t$ represents, which is the basis of the taxonomy in Section 5. A closely related architecture, the convolutional LSTM [34], replaces the matrix products inside the gates with convolutions so that the recurrent state keeps its spatial layout.

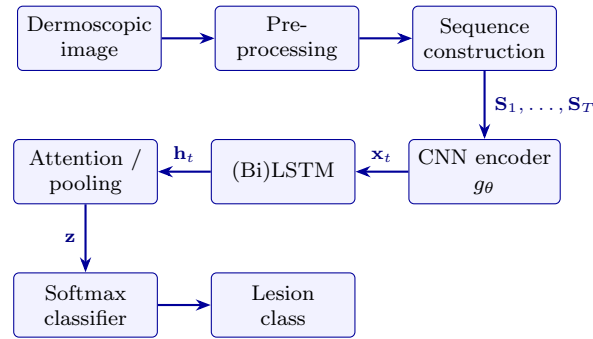


Fig. 1: Generic CNN–LSTM pipeline for skin lesion classification. The sequence $\mathbf{S}_1, \dots, \mathbf{S}_T$ may consist of patches, feature-map regions, multi-scale views or follow-up images

4 Datasets and Evaluation

4.1 Public Benchmarks

Table 1 summarises the datasets most often used by the reviewed studies. HAM10000 [42] and the ISIC challenge collections [7, 30] dominate. HAM10000 contains 10,015 dermoscopic images in seven diagnostic classes and is highly imbalanced, with melanocytic nevi forming roughly two thirds of the images. The ISIC 2020 dataset [30] is patient-centric, with multiple lesions per patient and contextual metadata, and has a melanoma prevalence below 2%. PH2 [23] is small but carefully annotated, and Derm7pt [19] pairs clinical and dermoscopic images with seven-point checklist labels, which makes it attractive for multimodal and multitask models. BCN20000 [8] adds difficult lesions from anatomically challenging sites.

4.2 Evaluation Metrics

Studies report accuracy, precision, recall (sensitivity), specificity, F1-score and the area under the receiver operating characteristic curve (AUC). Because of

Table 1: Public dermoscopy datasets used for CNN–LSTM skin cancer models

Dataset	Images	Classes	Notes
PH2 [23]	200	3	Nevi, atypical nevi, melanoma; expert segmentation masks
ISIC 2017 [7]	2,000 (train)	3	Melanoma, nevus, seborrhoeic keratosis
HAM10000 [42]	10,015	7	Multi-source; strongly imbalanced
Derm7pt [19]	1,011 cases	multi	Clinical + dermoscopic images, 7-point labels
BCN20000 [8]	19,424	8	Includes nail and mucosal lesions
ISIC 2019	25,331	8	Combines HAM10000, BCN20000 and others
ISIC 2020 [30]	33,126	2	Patient-level metadata; <2 % melanoma

class imbalance, overall accuracy can be dominated by the majority nevus class; balanced accuracy, per-class sensitivity for melanoma and AUC are more informative. Clinically, sensitivity for melanoma is usually prioritised, as a missed melanoma is far more costly than an unnecessary biopsy.

5 Taxonomy of CNN–LSTM Hybrids

The reviewed works can be organised along two axes: how the sequence fed to the recurrent stage is constructed, and how the recurrent stage is configured (Fig. 2).

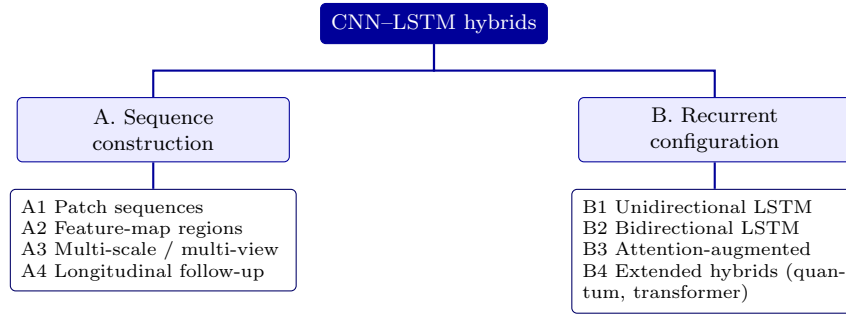


Fig. 2: Proposed taxonomy of CNN–LSTM hybrids for skin cancer detection

5.1 Sequence Construction

A1: Patch sequences. The image is divided into non-overlapping patches that are ordered (for example in raster order) and encoded by a time-distributed CNN; the LSTM then models dependencies across patches. Abohashish et al. [1] adopt this design, splitting each HAM10000 image into a patch sequence processed by time-distributed convolution and pooling layers combined with an LSTM, and

report an accuracy of 96.21 %. The strength of the approach is that it imposes explicit spatial reasoning; its weakness is that raster ordering is arbitrary, so the model must learn relations between patches whose sequence position does not reflect anatomical adjacency.

A2: Feature-map regions. Instead of raw patches, a pretrained backbone produces a deep feature map whose spatial positions, rows or channels are read as a sequence. Zareen et al. [45] combine an ImageNet-pretrained ResNet-50 with an LSTM for nine-class classification of ISIC archive images, and Rupakula et al. [31] feed features from ResNet50 and EfficientNet-B0 to LSTM layers on Derm7pt. This strategy reuses powerful transfer-learned representations and is computationally cheaper than encoding many patches, but it inherits the ordering problem at a coarser scale.

A3: Multi-scale and multi-view sequences. A sequence can also be formed from the same lesion at several magnifications, from crops centred on the lesion and its border, or from clinical and dermoscopic views of the same case, as available in Derm7pt [19]. The recurrent stage then acts as a learned fusion operator. This design matches the way dermatologists move between overview and detail, yet it remains less explored than A1 and A2.

A4: Longitudinal follow-up. The most natural use of a recurrent network is genuinely temporal data: sequential digital dermoscopy of the same lesion over months. Change over time is one of the strongest clinical indicators of melanoma, and a CNN-LSTM can in principle learn evolution patterns that no single image reveals. Public longitudinal datasets are, however, scarce, and the studies reviewed here operate almost exclusively on single images. This gap is discussed further in Section 8.

5.2 Recurrent Configuration

B1: Unidirectional LSTM. Most early hybrids stack one or two LSTM layers after the CNN and use the last hidden state for classification [28, 45].

B2: Bidirectional LSTM. Because spatial sequences have no inherent direction, BiLSTMs are a better fit. Hussein et al. [18] place a BiLSTM after a MobileNetV2 feature extractor and a simulated hybrid quantum layer for benign-malignant classification.

B3: Attention-augmented recurrence. Shaik et al. [33] combine a CNN with a BiLSTM and add spatial, channel and temporal attention. They compare GRU, LSTM and BiGRU variants with and without attention, and the attention-augmented CNN-BiLSTM gives the best result (92.73 % accuracy). Attention weights also offer a degree of interpretability by indicating which sequence elements drive the prediction.

B4: Extended hybrids. Recent work grafts recurrent stages onto other paradigms, including quantum-classical networks [18] and EfficientNet-B6

backbones proposed in a recent review [21]. In parallel, non-recurrent hybrids that combine CNNs with vision transformers [11, 22] or with handcrafted features [24] have become strong competitors, as has lesion-aware transformer modelling in related retinal imaging tasks [3].

6 Comparative Analysis

Table 2 compares representative studies. Values are reproduced as reported by the original authors; they were obtained on different datasets, class definitions and splits, and are therefore not directly comparable.

Table 2: Representative CNN-based and CNN–recurrent studies for skin cancer detection (metrics as reported by the original authors; NR = not reported in the abstract)

Study	Architecture	Recurrent unit	Data	Reported result
Esteva et al. 2017 [12]	Inception v3, transfer learning	None	≈130k clinical images	Dermatologist-level AUC
Haenssle et al. 2018 [15]	Inception v4	None	Dermoscopic reader study, 58 dermatologists	Outperformed most readers
Shukla et al. 2024 [35]	CNN + transfer learning + random forest	None	Two benign/malignant mole sets	Acc. up to 90.11 %
Naeem et al. 2024 [24]	Inception V3 + handcrafted features (SNC_Net)	None	ISIC 2019, 8 classes	Acc. 97.81 %
Mateen et al. 2024 [22]	U-Net + Inception-ResNet-v2 + ViT	None	ISIC 2020	Acc. 98.65 %
Zareen et al. 2024 [45]	ResNet-50 + LSTM	LSTM	ISIC archive, 9 classes (augmented)	Acc. 99.06 %
Abohashish et al. 2025 [1]	Patch sequence + time-distributed CNN	LSTM	HAM10000	Acc. 96.21 %
Shaik et al. 2025 [33]	CNN + spatial, channel and temporal attention	BiLSTM	Dermoscopic images	Acc. 92.73 %
Hussein et al. 2025 [18]	Hybrid quantum CNN + MobileNetV2	BiLSTM	ISIC-derived benign/malignant set	Test acc. 89.3 %
Rupakula et al. 2026 [31]	ResNet50 / EfficientNet-B0	LSTM	Derm7pt	Better than standalone CNNs (NR)
Ravikanth et al. 2026 [28]	CNN–LSTM (SkinWise)	LSTM	7 lesion classes	NR
Mahajan and Dashore 2025 [21]	EfficientNet-B6 + LSTM (concept)	LSTM	Review	Concept model

Several observations follow from Table 2. First, CNN–LSTM hybrids report accuracies between roughly 89 % and 99 %, which is in the same range as strong non-recurrent hybrids [22, 24]. The available evidence therefore does not show that the recurrent stage is uniformly superior; it shows that it is competitive and, in controlled ablations such as that of Shaik et al. [33], that bidirectional recurrence with attention outperforms simpler recurrent variants.

Second, the highest figures were obtained on balanced subsets created by augmentation [45] or on reduced class sets, whereas results on the full, imbalanced benchmarks are lower and harder to obtain. Third, careful reading of several primary studies reveals internal inconsistencies, such as differing values for the same metric or class counts that do not match the stated split, and some authors themselves report that differences between model variants are not statistically significant [18]. Such issues underline the need for standardised reporting before claims of superiority can be accepted.

Finally, few studies justify why their particular sequence ordering should help. Patch-raster and feature-row orderings are convenient but arbitrary, and the absence of ablations that shuffle the sequence order leaves open whether the LSTM is exploiting spatial context or simply acting as an additional non-linear aggregation layer.

7 Open Challenges

Class imbalance. Melanoma and several non-melanoma classes are heavily under-represented. Common remedies include weighted or focal losses [20], oversampling and augmentation; generative adversarial networks have been used to synthesise minority-class images in related cancer-imaging tasks [41]. Augmentation must be applied only after the train/test split, otherwise near-duplicates leak into the test set and inflate results.

Skin-tone diversity and generalisation. Public benchmarks are dominated by images of lighter skin. Daneshjou et al. [9] showed that dermatology AI systems lose performance on darker skin tones and uncommon diseases. External validation on independent cohorts remains rare among the reviewed hybrids.

Sequence justification. As noted above, most hybrids impose an artificial order on a static image. Without order-permutation ablations or comparison against equally sized non-recurrent heads, the contribution of the LSTM is difficult to isolate.

Interpretability. Clinicians need to know why a lesion is flagged. Gradient-based saliency methods such as Grad-CAM [32] and attention weights [33] provide partial explanations, but their faithfulness is seldom validated against dermoscopic criteria.

Robustness and security. Deep networks are vulnerable to small adversarial perturbations [27], and image-acquisition artefacts such as hair, ruler marks and colour-calibration differences can shift predictions. Robustness testing is largely absent from the reviewed works.

Computational cost. Time-distributed CNNs over many patches multiply inference cost, and recurrent layers are sequential by nature. This conflicts with deployment on smartphones and low-cost dermatoscopes in primary care.

Reproducibility. Code, trained weights and exact splits are often unavailable, and the inconsistencies discussed in Section 6 make independent verification difficult.

8 Future Directions

Longitudinal lesion monitoring. The strongest case for recurrent modelling is the analysis of image series acquired during follow-up. Curating public longitudinal dermoscopy datasets, with registered images of the same lesion over time, would let CNN–LSTM models learn clinically meaningful evolution.

Transformer–recurrent hybrids. Self-attention captures global context in parallel [11], whereas recurrence is efficient for long or streaming sequences. Combining a CNN or transformer encoder for spatial context with an LSTM for temporal follow-up is a promising division of labour.

Multimodal fusion. Age, sex, anatomical site and lesion history are informative [30]. A recurrent stage can fuse image features with metadata, or clinical with dermoscopic views [19], in a single learned representation.

Privacy-preserving training. Federated learning allows models to be trained across hospitals without sharing images [29], addressing both privacy regulation and the need for diverse training populations.

Lightweight deployment. Pruned and quantised backbones, GRU-based heads [6] and efficient patch selection can make hybrids practical on mobile devices at the point of care.

Standardised benchmarking. A common protocol that fixes patient-level splits, reports balanced accuracy, melanoma sensitivity and AUC with confidence intervals, and includes sequence-order ablations would make future comparisons meaningful.

9 Conclusion

This survey has reviewed hybrid CNN–LSTM models for skin cancer detection. CNNs remain the foundation of automated dermoscopy analysis, and adding an LSTM, especially a bidirectional and attention-augmented one, offers a principled way to integrate evidence across patches, regions, scales or time. Published hybrids report accuracies comparable to the best non-recurrent models on HAM10000, ISIC and Derm7pt. The evidence base, however, has clear limits: results are obtained under heterogeneous protocols, the benefit of an imposed sequence order on static images is seldom tested, and robustness, fairness across skin tones and interpretability are rarely evaluated. Progress will depend less on deeper architectures than on better data, particularly longitudinal and diverse cohorts, and on rigorous, reproducible evaluation. With those foundations, CNN–LSTM hybrids could become valuable decision-support tools for early melanoma detection, especially when applied to genuinely sequential follow-up imaging.

References

1. Abohashish, S.M.M., Amin, H.H., Elsedimy, E.I.: Enhanced melanoma and non-melanoma skin cancer classification using a hybrid LSTM-CNN model. *Scientific Reports* **15**, 24994 (2025). <https://doi.org/10.1038/s41598-025-08954-8>
2. Adegun, A., Viriri, S.: Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artificial Intelligence Review* **54**, 811–841 (2021). <https://doi.org/10.1007/s10462-020-09865-y>
3. Agarwal, A.K., Hamid, A.B.B.A., Ather, D., Tiwari, R.G., De, I., Krishna, K.R.: Lesion-aware ordinal transformer for diabetic retinopathy classification from fundus images. *Biomedical and Pharmacology Journal* **19**(1), 233–245 (2026). <https://doi.org/10.13005/bpj/3349>
4. Ather, D., Yusupov, I., Duggal, S., Kumar, R., Sagar, P., Jain, V.: Advancements in deep learning for early detection and diagnosis across multiple disease domains. In: *Proc. 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan. pp. 1460–1464. IEEE (2024). <https://doi.org/10.1109/ICTACS62700.2024.10840864>
5. Brinker, T.J., et al.: Skin cancer classification using convolutional neural networks: Systematic review. *Journal of Medical Internet Research* **20**(10), e11936 (2018). <https://doi.org/10.2196/11936>
6. Cho, K., et al.: Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 1724–1734 (2014). <https://doi.org/10.3115/v1/D14-1179>
7. Codella, N.C.F., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). In: *Proc. IEEE 15th International Symposium on Biomedical Imaging (ISBI)*. pp. 168–172. IEEE (2018). <https://doi.org/10.1109/ISBI.2018.8363547>
8. Combalia, M., et al.: BCN20000: Dermoscopic lesions in the wild. *arXiv:1908.02288* (2019)
9. Daneshjou, R., et al.: Disparities in dermatology AI performance on a diverse, curated clinical image set. *Science Advances* **8**(32), eabq6147 (2022). <https://doi.org/10.1126/sciadv.abq6147>
10. Donahue, J., et al.: Long-term recurrent convolutional networks for visual recognition and description. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2625–2634. IEEE (2015). <https://doi.org/10.1109/CVPR.2015.7298878>
11. Dosovitskiy, A., et al.: An image is worth 16×16 words: Transformers for image recognition at scale. In: *Proc. International Conference on Learning Representations (ICLR)* (2021). *arXiv:2010.11929*
12. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017). <https://doi.org/10.1038/nature21056>
13. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* **18**(5–6), 602–610 (2005). <https://doi.org/10.1016/j.neunet.2005.06.042>
14. Gupta, S., Abdul Hamid, A.B., Nyamasvisva, T.E., Jain, V., Tyagi, N., Mun, N.K., Ather, D.: Cotton disease prediction using deep transfer learning: Comparative analysis of ResNet50, VGG16 and InceptionV3 models. *International Journal of*

- Engineering, Science and Information Technology **5**(3), 588–594 (2025). <https://doi.org/10.52088/ijesty.v5i3.1116>
15. Haenssle, H.A., et al.: Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology* **29**(8), 1836–1842 (2018). <https://doi.org/10.1093/annonc/mdy166>
16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778. IEEE (2016). <https://doi.org/10.1109/CVPR.2016.90>
17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Computation* **9**(8), 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
18. Hussein, A.A., Montaser, A.M., Elsayed, H.A.: Skin cancer image classification using hybrid quantum deep learning model with BiLSTM and MobileNetV2. *Quantum Machine Intelligence* **7**(2), 66 (2025). <https://doi.org/10.1007/s42484-025-00288-y>
19. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE Journal of Biomedical and Health Informatics* **23**(2), 538–546 (2019). <https://doi.org/10.1109/JBHI.2018.2824327>
20. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *Proc. IEEE International Conference on Computer Vision (ICCV)*. pp. 2980–2988. IEEE (2017). <https://doi.org/10.1109/ICCV.2017.324>
21. Mahajan, S.T., Dashore, P.: Hybrid deep learning models for skin lesion classification: A comparative review and future directions. *EPJ Web of Conferences* **341**, 01051 (2025). <https://doi.org/10.1051/epjconf/202534101051>
22. Mateen, M., Hayat, S., Arshad, F., Gu, Y.-H., Al-antari, M.A.: Hybrid deep learning framework for melanoma diagnosis using dermoscopic medical images. *Diagnostics* **14**(19), 2242 (2024). <https://doi.org/10.3390/diagnostics14192242>
23. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R.S., Rozeira, J.: PH2 – A dermoscopic image database for research and benchmarking. In: *Proc. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. pp. 5437–5440. IEEE (2013). <https://doi.org/10.1109/EMBC.2013.6610779>
24. Naeem, A., Anees, T., Khalil, M., Zahra, K., Naqvi, R.A., Lee, S.-W.: SNC_Net: Skin cancer detection by integrating handcrafted and deep learning-based features using dermoscopy images. *Mathematics* **12**(7), 1030 (2024). <https://doi.org/10.3390/math12071030>
25. Naitik, Raj, D., Rajan, D.K., Gupta, A.K., Agrawal, A.K., Krishna, K.R.: Enhancing toxic comment detection with BiLSTM-based deep learning model. In: *Proc. International Conference on Information Science and Communications Technologies (ICISCT)*, Seoul, Korea. pp. 206–211. IEEE (2024). <https://doi.org/10.1109/ICISCT64202.2024.10956568>
26. Pathak, S., Quraishi, S.J., Singh, A., Singh, M., Arora, K., Ather, D.: A comparative analysis of machine learning models: SVM, naïve Bayes, random forest, and LSTM in predictive analytics. In: *Proc. 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*. pp. 790–795. IEEE (2023). <https://doi.org/10.1109/ICTACS59847.2023.10390255>
27. Pranjal, Vaishnavi, Divyansh, Raj, D., Jain, V., Agarwal, A.K.: Adversarial attacks on neural networks. In: *Cyber Intelligence and Information Retrieval (CIIR 2023)*. LNNS, vol. 1025. Springer, Singapore (2024). <https://doi.org/10.1007/>

- 978-981-97-3594-5_34
28. Ravikanth, V., Jahnavi, D., Hemanth, V.R., Reddy, J.J., Kumar, D.V.H.: Skin-Wise: Hybrid CNN-LSTM for cancer diagnosis. In: Evolution in Signal Processing and Telecommunication Networks (ICMEET 2025). LNEE, vol. 1601, pp. 205–211. Springer, Cham (2026). https://doi.org/10.1007/978-3-032-20544-5_18
 29. Rieke, N., et al.: The future of digital health with federated learning. npj Digital Medicine **3**, 119 (2020). <https://doi.org/10.1038/s41746-020-00323-1>
 30. Rotemberg, V., et al.: A patient-centric dataset of images and metadata for identifying melanomas using clinical context. Scientific Data **8**, 34 (2021). <https://doi.org/10.1038/s41597-021-00815-z>
 31. Rupakula, T., Murugadoss, R., Karthikeyan, A., Aruna, M., Periyasamy, P.R., Krishnan, K.: Hybrid intelligence system CNN-LSTM architecture for accurate skin cancer classification using dermoscopic image analysis. Genetics and Molecular Research **25**(9s) (2026). <https://doi.org/10.4238/c4mb4z83>
 32. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proc. IEEE International Conference on Computer Vision (ICCV). pp. 618–626. IEEE (2017). <https://doi.org/10.1109/ICCV.2017.74>
 33. Shaik, A., Dutta, S.S., Sawant, I.M., Kumar, S., Balasundaram, A., De, K.: An attention based hybrid approach using CNN and BiLSTM for improved skin lesion classification. Scientific Reports **15** (2025). <https://doi.org/10.1038/s41598-025-00025-2>
 34. Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-C.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 28, pp. 802–810 (2015)
 35. Shukla, M.M., Tripathi, B.K., Dwivedi, T., Tripathi, A., Chaurasia, B.K.: A hybrid CNN with transfer learning for skin cancer disease detection. Medical & Biological Engineering & Computing **62**(10), 3057–3071 (2024). <https://doi.org/10.1007/s11517-024-03115-x>
 36. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proc. International Conference on Learning Representations (ICLR) (2015). arXiv:1409.1556
 37. Singh, S., Prakash, P., Baghel, G., Singh, A., Raj, D., Agrawal, A.K.: Banana crop health: A deep learning-based model for disease detection and classification. In: Proc. 27th International Symposium on Wireless Personal Multimedia Communications (WPMC), Greater Noida, India. pp. 1–6. IEEE (2024). <https://doi.org/10.1109/WPMC63271.2024.10863138>
 38. Singhal, P., Kaushik, M., Mansoori, A., Agrawal, A.K., Raj, D.: Transformer-assisted deep learning model for medical image enhancement. In: Proc. 13th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India. pp. 1–7. IEEE (2026). <https://doi.org/10.23919/INDIACom70271.2026.11526310>
 39. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **71**(3), 209–249 (2021). <https://doi.org/10.3322/caac.21660>
 40. Tan, M., Le, Q.V.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proc. 36th International Conference on Machine Learning (ICML). PMLR, vol. 97, pp. 6105–6114 (2019)

41. Tripathi, R.P., Khatri, S.K., Van Greunen, D., Ather, D.: Enhancing breast cancer diagnosis through segmentation-driven generative adversarial networks for synthetic mammogram generation. In: Proc. 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS). pp. 1078–1082. IEEE (2023). <https://doi.org/10.1109/ICTACS59847.2023.10389935>
42. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data* **5**, 180161 (2018). <https://doi.org/10.1038/sdata.2018.161>
43. Walia, S., Singh, S., Mor, J., Raj, D., Jain, V., Krishna, K.R.: An efficient Caffe model based multiple object detection and recognition system using R-CNN. In: *Intelligent Computing and Communication Techniques*, pp. 678–685. CRC Press (2025)
44. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.-A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Transactions on Medical Imaging* **36**(4), 994–1004 (2017). <https://doi.org/10.1109/TMI.2016.2642839>
45. Zareen, S.S., Sun, G., Kundi, M., Qadri, S.F., Qadri, S.: Enhancing skin cancer diagnosis with deep learning: A hybrid CNN-RNN approach. *Computers, Materials & Continua* **79**(1), 1497–1519 (2024). <https://doi.org/10.32604/cmc.2024.047418>