| **Model** | **Architecture Type** | **Dataset Used** | **Metric Used** | **Reported Accuracy / Score** | **Reference** |
| --- | --- | --- | --- | --- | --- |
| **U-Net** | CNN | ISIC 2018 | Dice Coefficient | 85.2% | [Ali et al., 2022](https://doi.org/10.1007/s11042-021-11708-z) |
| **Attention U-Net** | CNN + Attention | ISIC 2018 | Dice Score | 86.5% | [Jabbar & Khan, 2022](https://doi.org/10.1007/978-981-19-3391-2_25) |
| **TransUNet** | CNN + ViT Hybrid | ISIC 2018, 2017 | Dice Score | 88.1% | [Chen et al., 2021](https://arxiv.org/abs/2102.04306) |
| **Swin-Unet** | Pure Transformer | ISIC 2018, Synapse | Dice Score | 88.9% | [Cao et al., 2021](https://arxiv.org/abs/2105.05537) |
| **MedT** | Pure Transformer | ISIC 2018 | Dice Score | 87.8% | [Valanarasu et al., 2021](https://arxiv.org/abs/2102.10662) |
| **GS-TransUNet** | ViT + Gaussian Splatting | ISIC 2018 | Dice Score | 89.3% | [Zhao et al., 2024](https://arxiv.org/abs/2502.16748) |
| **ScaleFusionNet** | Swin Transformer + DeformConv | ISIC 2018 | Dice Score | **89.5%** | [Shao et al., 2024](https://arxiv.org/abs/2503.03327) |
| **SUTrans-NET** | Dual Encoder (CNN + Transformer) | ISIC 2018 | Accuracy | 90.1% | [Wang et al., 2023](https://doi.org/10.3390/s23062936) |
| **Attention Swin U-Net** | Swin Transformer + Cross-Attn | ISIC 2018 | Dice Score | **90.6%** | [Kim et al., 2024](https://arxiv.org/abs/2210.16898) |
| **ViT-UNet (Our Project)** | CNN + ViT Hybrid | ISIC 2018 (assumed) | Dice Score / Accuracy | ~**91.2%** *(as reported)* | *Our Project* |
|  | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model / Study** | **Architecture Type** | **Dataset(s)** | **Key Features** | **Performance Metrics** | **Reference** |
| **Hybrid Deep Learning Framework** | U-Net + Inception-ResNet-v2 + ViT | ISIC 2020, HAM10000 | Combines U-Net for segmentation, Inception-ResNet-v2 for feature extraction, and Vision Transformer for feature refinement | Accuracy: 98.65%; Sensitivity: 99.20%; Specificity: 98.03% | [PubMed, 2024](https://pubmed.ncbi.nlm.nih.gov/39410645/) |
| **Advanced Deep Learning Models** | Context Aggregation-based DNN | ISIC 2020 | Utilizes morphological operations and context aggregation for preprocessing and segmentation | Classification Accuracy: 93.40% | [MDPI Sensors, 2025](https://www.mdpi.com/1424-8220/25/3/594) |
| **Skin Lesion Classification via Ensemble Method** | Modified Inception ResNet v2 + EfficientNet-B4 | ISIC 2018 | Incorporates Soft-Attention mechanism to enhance feature extraction | Specific performance metrics not provided | [Springer, 2024](https://rd.springer.com/article/10.1007/s11042-024-19837-x) |
| **Boundary-Aware Segmentation Network (BASNet)** | CNN + Transformer | ISIC 2016/2017/2018, HAM10000 | Applies hybrid loss and residual refinements to handle occlusion and poor contrast images | Specific performance metrics not provided | [Frontiers in Medicine, 2025](https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2025.1524146/full) |
| **Skin Lesion Segmentation Model Based on Improved U2Net** | Improved U2Net | ISIC 2018 | Enhances U2Net architecture for better segmentation performance | Specific performance metrics not provided | [ACM ISAIMS, 2023](https://dl.acm.org/doi/10.1145/3644116.3644233) |
| **Skin Cancer Segmentation and Classification Using ViT** | Vision Transformer | HAM10000 | Utilizes pre-trained ViT for segmentation and classification | Classification Accuracy: 96.15% | [ResearchGate, 2024](https://www.researchgate.net/publication/377955561_Skin_Cancer_Segmentation_and_Classification_Using_Vision_Transformer_for_Automatic_Analysis_in_Dermatoscopy-Based_Noninvasive_Digital_System) |
| **Deep Learning-based Skin Lesion Segmentation and Classification** | U-Net + CNN + ViT | HAM10000 | Employs U-Net for segmentation followed by CNN and ViT for classification | Specific performance metrics not provided | [ResearchGate, 2024](https://www.researchgate.net/publication/382398480_Deep_Learning-based_Skin_Lesion_Segmentation_and_Classification) |
| **Health of Things Melanoma Detection System** | Deep Learning + Fine-Tuning Models | Not specified | Applies deep learning and fine-tuning models in edge computing environments | Detection Accuracy: 96.39%; Segmentation Accuracy: 96.50% | [Frontiers in Communications and Networks, 2024](https://www.frontiersin.org/journals/communications-and-networks/articles/10.3389/frcmn.2024.1376191/full) |