SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Comparative Analysis of CNN And YOLOv5 For Automated Skin Burn Severity Classification

Dr. Sunil Karamchandani[1], Aryan Satam[2], Vansh Tank[3], Siddhant Sanghoi[4], Shrirang S. Deshmukh[5], Dr. Dhruvi Shah[6]

[1]Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India.
E-mail: sunil.karamchandani@djsce.ac.in
[2]Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India. E-mail: aryansatam22@gmail.com
[3]Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India. E-mail: vanshtank999@gmail.com
[4]Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India. E-mail: siddhantsanghoi@gmail.com
[5]Electronics and Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India.
E-mail: hrirangdeshm1313@gmail.com
[6]HBT Medical College and Dr RN Cooper Municipal General Hospital, Mumbai, India. E-mail: dhruvishah401@gmail.com

ABSTRACT
The main objective of this study is to improve burn depth evaluation accuracy through automated deep learning models including Convolutional Neural Networks (CNNs) and YOLOv5. The standard medical approach for evaluating burns involves subjective assessments which generate unpredictable variations between observers thus resulting in incorrect diagnoses. We developed an automated system that provides immediate precise classification of burn wounds into first, second and third-degree categories. The research evaluates the performance of CNN-based models together with YOLOv5 by measuring their detection accuracy and processing speed and operational efficiency. Our research shows YOLOv5's single-stage detection system delivers better performance in accuracy and real-time detection compared to CNNs. The study demonstrates how automated burn wound evaluation can be enhanced for clinical practice which will decrease misdiagnosis rates while enhancing patient care.

Keywords: Terms—skin burns, deep learning, CNN, YOLOv5, medical image analysis, burn classification

Introduction

The worldwide healthcare system faces burn injuries as a significant public health issue which leads to more than one million moderate-to-severe burn cases in India every year. Approximately a quarter of all burn cases affect children younger than 15 years while burns rank among the top causes of fatal injuries for children between 1 to 9 years old. The depth of a burn determines its severity which directly affects both treatment approaches and patient recovery outcomes. The accurate identification of burn severity needs to happen quickly because it enables proper medical treatment while reducing both infection risk and scarring and death rates. Burn severity assessment through current clinical methods depends mainly on human observation of burns and professional judgment but this approach produces incorrect results which delays medical intervention. Modern machine learning and deep learning technologies present an opportunity to advance healthcare by delivering automated quick assessments of burn severity. The research study evaluates how well Convolutional Neural Networks (CNNs) perform against YOLOv5 for solving this important healthcare issue.

Challenges with Traditional Burn Assessment:
1) Current clinical approaches face multiple significant problems in their implementation
2) Significant variations exist in how different medical professionals classify burn injury depth
3) Medical professionals with expert skills must perform these assessments
4) The availability of advanced imaging tools like Laser Doppler Imaging (LDI) remains restricted
5) Medical practitioners require immediate assessment methods that are both fast and objective

Machine Learning in Burn Diagnosis:
Deep learning together with machine learning has revolutionized medical image analysis through automatic feature extraction and classification of complex visual data. The powerful nature of CNNs for image classification stems from their capability to develop feature hierarchies starting from edges to textures and shapes without requiring manual feature extraction. CNN-based burn severity classification methods have achieved good results but fail to deliver the necessary real-time performance needed in emergency care settings. YOLOv5 represents a state-of-the-art single-stage object detection model which combines detection and classification processes into one efficient operation. YOLOv5 delivers fast burn classification together with image-based burn localization in real-time. The

implementation of these deep learning models in clinical settings will boost diagnostic precision while decreasing evaluation durations and enabling scalable burn assessment solutions which are perfect for emergencies.

Literature Survey

Deep learning techniques have grown increasingly influential in the analysis of medical images, such as in the assessment of burn severity. Abubakar et al. (2020) showed that deep learning features from CNNs significantly improve the accuracy of burn depth diagnosis over traditional clinical techniques, therefore setting a standard for automatic diagnosis systems in burns management. From this developed work, Rahman et al. (2024) compare several deep learning models in detail with diagnostics versus computational efficiency, an important consideration for real-world clinical use. Moura et al. (2021) presented an extensive systematic review and emphasized how artificial intelligence has been playing a critical transformative role in the assessment of burn injuries and their treatment, as this illustrates better clinical outcomes along with workflow optimization.

Core to these developments are pioneering CNN architectures, such as those by Krizhevsky et al. (2017), whose contribution to ImageNet classification sealed the status of deep CNNs as powerful image recognition techniques across domains including healthcare. The YOLO family of models (You Only Look Once), especially the works known as YOLOv3 by Redmon and Farhadi (2018) and YOLOv5 by Jocher et al. (2021), enabling real-time object detection with considerable accuracy, is particularly suitable for tasks in medical imaging that require the identification and categorization of anomalies such as burn wounds with speed. He et al. (2016) proposed residual networks (ResNet), which mitigate some of the problems that accompany the training of deep models and thus enable deeper models for medical image interpretation. On this basis, the EfficientNet architecture developed by Tan & Le (2019) optimized model scaling further; with higher accuracy, using fewer parameters became important for the deployment of models in clinically resource-constrained environments.

Applications of these models in healthcare go beyond burn severity classification. Sharma et al. (2024) applied detection based on YOLO to recognize complex symbols, demonstrating the flexibility of the architecture that can potentially be translated to recognizing intricate patterns of burn wounds. Transfer learning strategies utilizing VGG16 have been successfully used for lung opacity classification by Singla et al. (2024) and Bhuria et al. (2024), showing how pre-trained CNNs can easily be repurposed for specialized medical image classification tasks with limited datasets. Naman et al. (2024) further demonstrated the applicability of CNNs in biomedical classification through the identification of herbal drugs, further reinforcing the versatility of deep learning in the area of healthcare-related image classification.

Improvements in variants of YOLO intended for specific detection applications, such as small object detection by Carrasco et al. using T-YOLO, yield methods that might be translated to accurate detection and classification of burn wounds with highly variable lesion sizes. Likewise, Xu et al. used YOLOv8s in early plant detection in complex environments, similar to the occlusion and lighting variability challenges faced in many clinical imaging conditions. Improvements to the YOLO architecture like the YOLO-C3GTR model developed by Usman et al. offer higher accuracy in detection, especially architectural innovations that are likely to benefit further detailed medical image analysis.

Reviews, such as by Sakib et al. (2018), provide basic understanding regarding CNN architectures and their broad applications, which was important for the exploration of new models on burn severity assessment. Works like those of Jeschke et al. and Siu et al. introduce the clinical aspect and developments around burn wound healing, proposing real-world scenarios for deploying automatic diagnosis systems. In all, this collective contributed to the continuous development of burn severity classification using AI with high accuracy, efficiency, and scalability and melded strong architectures with real clinical-world considerations to improve diagnostic precision and patient care.

Cnn Architecture

Figure 1 presents the schematic architecture of the Convolutional Neural Network (CNN) framework utilized for automatic burn severity classification. The sequential design enables progressive extraction of hierarchical image features, which are subsequently mapped to diagnostic categories through dense layers.

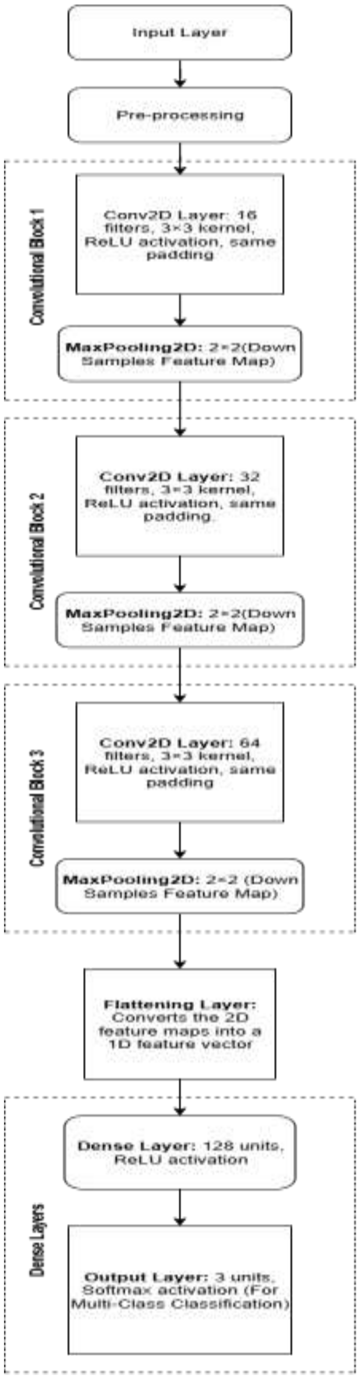

Fig 1: CNN Custom Architecture

Layered Architecture Description:

The Input Layer: Inputs 128×128×3 RGB images.

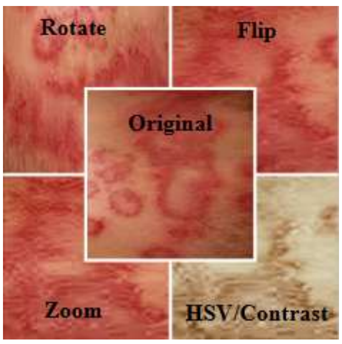

Fig 2: Dataset Image

Fig 3: Preprocessed dataset image

Pre-processing:
Pre-processing is a crucial step that ensures all input images fed to the network are in a consistent and suitable form for learning. The pre-processing starts with resizing, making all images of uniform dimensions for batch processing and easy alignment with the expected neural network input shape. Following that is normalization by scaling pixel values between 0 and 1. Normalization also accelerates model convergence and makes training very stable with the avoidance of large updates when gradients propagate inside the neural network. Moreover, augmentation artificially increases the number of images in a training dataset and encourages model robustness. Augmentations may include random rotations, flips, zooms, or other non-geometric changes like brightness or contrast adjustment. With the introduction of such transformations, a wider variety of examples will be seen by the model, hence it will have better generalization for unseen clinical images, while avoiding overfitting.

Convolutional Blocks:
Pre-processed images enter a sequence of convolutional blocks, forming the main units for feature extraction within the network. Each block is typically made of stacked Conv2D layers. The image is convolved with 3×3 convolutional kernels, extracting local features at an increasing level of abstraction. These layers introduce non-linearity with the ReLU activation function for assisting the network in learning complex mappings between image inputs and classification targets. Convolutions are performed with consistent padding to ensure that feature maps retain their original spatial dimensions after convolution - it can be an important property for tasks requiring fine spatial localization.
These blocks, composed of convolution and pooling operations in a 2×2 window, are interleaved to progressively down-sample spatial dimensions, hence summarizing feature-rich regions in a way that significantly reduces computational demands. Notice that the depth of the convolutional filters increases across successive blocks: e.g., 16, 32, 64; this better prepares the model for recognizing intricate visual patterns and textural cues which are characteristic of different classes of burn severity.

Flattening Layer:
In the process of flattening, which follows convolution, the multi-dimensional feature maps created by previous layers are reduced to a single one-dimensional vector. This representation consolidates the spatially distributed features into a form suitable for interpretation by dense, fully connected layers. Flattening thus acts as a bridge, allowing seamless transition from spatial pattern recognition to holistic decision-making. Dense Layers: The dense layers handle the classification stage. First, a fully connected hidden layer with 128 ReLU units further distills the information contained in the feature vector in order to capture the subtle and non-linear relationships that are important for the estimation of burn severity. The last output layer is a dense layer with three softmax units, one for each class of burn severity: first, second, or third degree. By using the softmax activation, class probabilities are obtained such that the model not only provides a classification but also gives a measure of its confidence in every prediction, which is really valuable in clinical decision support systems.

CNN-BASED APPROACH

Shallow CNN:
It is built from two blocks of convolution (64 and 32 filters) in which every block is made up of dual Conv2D layers, batch normalization, max pooling, and static dropout (0.25/0.3). To this structure, one dense layer (128 units, 0.3 dropout) is added, finished with the softmax output employed in three-class classification of burns.

Deep CNN:
It is composed of four consecutive filter-increasing convolutional blocks in order (32, 64, 128, 256) with dropout rates (0.25–0.5), all with max pooling and normalization per batch. We then have two dense (256 then 128 units, 0.5/0.3 dropout) before output in the form of softmax.

Burn Severity Classification Algorithm:
BEGIN
// Image Pre-processing
preprocessed_image ← PreprocessImage(image)
IF model_type = 'shallow' THEN
// Block 1
x ← 2×Conv2D(32) + BatchNorm + MaxPooling + Dropout(0.25)
// Block 2
x ← 2×Conv2D(64) + BatchNorm + MaxPooling + Dropout(0.3)
// Global Average Pooling
x ← GlobalAveragePooling(x)
// Dense + Dropout
x ← Dense(128) + Dropout(0.3)
ELSE IF model_type = 'deep' THEN
// Block 1

```
x ← 2×Conv2D(32) + BatchNorm + MaxPooling + Dropout(0.25)
// Block 2
x ← 2×Conv2D(64) + BatchNorm + MaxPooling + Dropout(0.3)
// Block 3
x ← 2×Conv2D(128) + BatchNorm + MaxPooling + Dropout(0.4)
// Block 4
x ← 2×Conv2D(256) + BatchNorm + MaxPooling + Dropout(0.5)
// Global Average Pooling
x ← GlobalAveragePooling(x)
// Dense + Dropout
x ← Dense(256) + Dropout(0.5)
x ← Dense(128) + Dropout(0.3)

ENDIF
// Softmax Output
output ← Softmax(x, 3 classes)
// Burn Severity Classification
severity ← ArgMax(output)   // 1st, 2nd, or 3rd degree
// Performance Analysis (handled elsewhere)
```

RETURN severity
END

Both the CNNs were trained using the Adam optimizer, with categorical cross-entropy loss as the objective function for the 3 classes. The initial learning rate was set to 0.001 for the shallow model 0.0001 for the deeper variant, with early stopping based on validation F1 score to prevent overfitting. A consistent batch size of 32 was used in all experiments. All experiments were performed on an NVIDIA H100 PCIe GPU with 80 GB memory.

Table 1: Overview of CNN models

Comparative Overview	Shallow CNN	Denser CNN
Depth	2 convolutional blocks	4 convolutional blocks
Filter Progression	32 □64	32 □ 64□128□256
Dense Layers	1 Dense (128) + GAP	2 Dense (256), Dense (128) + GAP
Dropout Strategy	Fixed: 0.25 after convolutional blocks; 0.3 after dense layers	Progressive: 0.25□0.3□0.4□0.5
Activation	ReLU	ReLU
Epochs	50	100
Batch Normalisation	After each conv2D	After each conv2D
Inference Speed	10-40 (NVIDIA H100 80GB)	5-15 (NVIDIA H100 80GB)
Validation Accuracy (%)	66	76
Initial Learning Rate	1E-03(0.001)	1E-04(0.0001)
Observed Constraints	Limited ability to capture complex burn textures	Less suitable for edge deployment

Observation:
The CNN-deep model obtained higher accuracy (76% vs. 66%) and superior F1 values, thus being suitable for the subtle classification of burns. But it required significantly more computing power and ran more slowly in inference. The light CNN provided faster and less complex inference, with lower accuracy, which was best for resource-constrained situations in real-time use cases.

Yolov5 Architecture

The YOLOv5 model leverages a purpose-built, single-stage detection pipeline tailored to the clinical challenges of burn wound analysis. Unlike generic CNN classifiers, it performs localization, grading, and quantification simultaneously—capabilities that are critical for triage, %TBSA calculation, surgical decision-making, and longitudinal wound monitoring in real-world care.

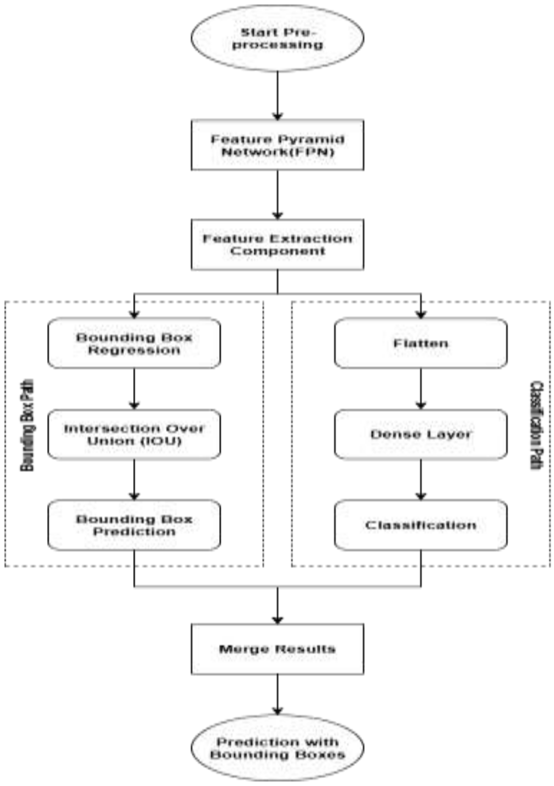

Fig. 4. YOLOv5 Architecture

Workflow & Core Components:

Pre-processing: Images are resized and normalized to minimize variability caused by lighting and complexion in order to obtain a strong input baseline for burn detection.

Backbone (CSP, 53 Layers): The nuanced burn features, such as irregular texture, jagged contours, and gradual transitions in color, are captured in the deep CSP-based backbone. This allows for consistent discrimination between superficial, deep, and mixed-thickness burns, which is paramount for clinical level determination as well as for surgical intent.

Feature Pyramid Network (FPN): FPN guarantees strong detection for minor and large-scale burns, pulling out feature maps at a series of multiple spatial resolutions, the same as clinicians assess lesions at a range of sizes and depths in practice.

Spatial Pyramid Pooling (SPP): SPP preserves core local information while enriching context so that focal as well as diffuse burn injuries are detected in a single image.

Dual Prediction Heads:

• Bounding Box Regression with GIoU loss: Creates narrow, shape-aware contours around wounds, on which accurate %TBSA estimations for fluid administration and prediction are based.

• Lesion-Based Classification: Suggests a region-level burn severity grade allocation for its use in verifying objective, reproducible wound severity assessments for all lesions identified.

• Result Combination: All outputs are combined into structured per-wound predictions: each lesion is assigned a mapped location, severity grade, and confidence value—all in standard clinical notation and optimally complementary for team-oriented care.

Table 2: Clinically Critical Features:

Feature	Clinical Impact
Depth (53 Layers)	Models nuanced texture, edge definition, and color for precise grading.
SPP	Enables full-spectrum assessment (from large burns to small lesions).
GIoU Loss	Refines wound perimeter accuracy; essential for quantifying burn area for therapy.

Feature	Clinical Impact
Multi-head Output	Provides wound-level location, severity, and confidence-supporting triage, graft sizing, and longitudinal monitoring

• YOLOv5: Provides, for each wound, a bounding box (location), burn grade (severity), and confidence score.
• CNNs: Output a single disease class for the entire image, with no spatial information.
• Clinical implication: While a CNN can state "this image has a deep burn," only YOLOv5 can deliver "three lesions: one superficial, two deep, spanning 18% TBSA." Such explicit, per-lesion outputs are crucial for fluid calculation, operative planning, and tracking healing or complications over time.

Table 3: Comparative Model Architecture

Aspect	Custom (Deep CNN)	YOLOv5
Architecture Type	Sequential Classifier	End-to-End Detector
Backbone Depth	4 Convolutional Blocks	53-layers CSP
Pooling Strategy	Global Average Pooling	Spatial Pyramid Pooling
Detection Head	None	Anchor-based regression+ Classification
Loss Function	Cross-Entropy	GIoU + Class Loss
Output	Single burn grade per image	Per-lesion: Location + Grade + Confidence

Table 4: Training Setup:

Hyperparameter	Custom (Deep CNN)	YOLOv5
Epochs	100	200
Optimizer	Adam	SGD
Batch Size	32	16
Data Augmentation	Rotation, Flip, Contrast, Zoom	HSV, Scaling

Yolov5 Advantages Over CNN (Comparative Analysis)

Table 5: Performance metrics of the CNN1 model

Degree Of Burn	Samples	Support	CNN1 Precision	CNN1 Recall	CNN1 F1 Score
1st Degree	5631	285	0.69	0.76	0.73
2nd Degree	6377	290	0.64	0.57	0.6
3rd Degree	2216	113	0.61	0.65	0.63
Average	-	-	0.64667	0.66	0.653

Table 6: Performance metrics of the CNN2 model

Degree Of Burn	Samples	Support	CNN2 Precision	CNN2 Recall	CNN2 F1 Score
1st Degree	5631	285	0.74	0.82	0.78
2nd Degree	6377	290	0.75	0.72	0.73

Degree Of Burn	Samples	Support	CNN2 Precision	CNN2 Recall	CNN2 F1 Score
3rd Degree	2216	113	0.85	0.7	0.77
Average	-	-	0.78	0.746	0.76

Tables 5 and 6 present the performance measures taken for two different CNN models, CNN1 and CNN2, against different classes of burns. Both architectures are able to distinguish between three classes of burn severity, though with a difference in performance. Whereas CNN1 has an average precision of 0.65, recall of 0.66, and F1-score of 0.65, depicting moderately good but consistent performance, CNN2 presents an improvement with average metrics at 0.78 for precision, 0.75 for recall, and 0.76 for the F1-score. In particular, the increased precision and recall in all classes in CNN2 indicate enhanced discriminatory power and reliability, especially towards complex cases like third-degree burns. It is also noted from the data that both frameworks perform comparatively well on first-degree cases and poorly on third-degree cases—a trend important for future refinement in automated burn classification pipelines.

Table 7: Performance metrics of the YOLOV5(Nano & Small)

Degree of Burn	Samples	Support	YOLOv5 Nano (Precision)	YOLOv5 Nano (Recall)	YOLOv5 Nano (F1 Score)	YOLOv5 Small (Precision)	YOLOv5 Small (Recall)	YOLOv5 Small (F1 Score)
1st Degree	5631	448	0.816	0.703	0.755	0.917	0.759	0.830
2nd Degree	6377	448	0.727	0.745	0.735	0.751	0.815	0.781
3rd Degree	2216	448	0.808	0.651	0.721	0.92	0.722	0.80
Average		-	0.783	0.699	0.731	0.862	0.765	0.767

Table 8: Performance metrics of the YOLOV5(Medium & XL)

Degree of Burn	Samples	Support	YOLOv5 Medium (Precision)	YOLOv5 Medium (Recall)	YOLOv5 Medium (F1 Score)	YOLOv5 XL(Precision)	YOLOv5 XL (Recall)	YOLOv5 XL (F1 Score)
1st Degree	5631	448	0.929	0.78	0.848	0.929	0.78	0.848
2nd Degree	6377	448	0.754	0.745	0.749	0.754	0.745	0.749
3rd Degree	2216	448	0.913	0.727	0.809	0.913	0.727	0.809
Average		-	0.865	0.750	0.548	0.865	0.750	0.781

The models are evaluated based on 1st-degree, 2nd-degree, and 3rd-degree burns with the sample sizes and support for each class given.

Metrics Reported: Precision, recall, and F1 score are provided for all models and classes to facilitate granular as well as overall comparison of performance.

Model Coverage:
CNN 1 and CNN 2 demonstrate balanced performance, with F1 measures from 0.65–0.78 for individual classes and mean F1 measures for all classes of 0.65 (CNN 1) and 0.76 (CNN 2) accordingly.
The YOLOv5 models are always superior to CNNs in all metrics as well as all categories, with very high recall and precision, as evidenced in F1 values for the top-performing in YOLOv5 (Medium and XL) as 0.848 for 1st-degree burns, 0.749 for 2nd-degree, and 0.809 for 3rd-degree. Average F1 in YOLOv5 XL is 0.781, markedly higher than in

both CNN variants.

YOLOv5 Advantage: All variants of YOLOv5 are better than their CNN counterparts, particularly in terms of F1 score, which.
Consistency: YOLOv5 models show more balanced performance across all burn classes, particularly excelling in difficult-to-classify cases like 3rd-degree burns.

Clinical Relevance: Increased accuracy, recall, and F1 value obtained in YOLOv5 models suggest their promising potential for more reliable, real-time clinical burn grading from standard CNN-based methods, enabling improved diagnostic accuracy and reproducibility.

Yolov5 Dataset And Augmentation

All models were trained and evaluated on the SVKM AI/ML high-performance computing (HPC) cluster. The experiments utilized a compute node equipped with two NVIDIA H100 PCIe GPUs, each featuring 80 GB of GPU memory. The node was configured with NVIDIA driver version 535.86.10 and CUDA 12.2. The hardware setup also included dual Intel Xeon Gold 6438Y+ processors and 512 GB of DDR5 RAM. Job management for training and inference tasks was handled through the Altair Access job submission portal.
The research works with a carefully curated dataset of 15,000 background and burn wound images borrowed from the public repository Roboflow. The dataset covers the whole range of severities, from first to third-degree burns to normal (background) skin, with sample distribution details as given below.

Table 9: Dataset Summary

Attribute	**Description**
Total Images	15,000
Class Samples	1st-Degree: 5631 2nd-Degree:6377 3rd-Degree:2216 Background:776
CNN Input Size	128x128 pixels, RGB
YOLOv5 Input Size	640x640 pixels, RGB
Source	Roboflow (public open dataset)
Annotation	Manual: Class (CNN), bounding box + class (YOLOv5)
Augmentation	Rotation, flip, contrast, zoom, HSV, Scaling

Data Annotation

In CNN-based image classification, images in the dataset are annotated with a single label representing the general grade of burn severity, namely, first-degree, second-degree, third-degree, or background. This allows the neural network to be sensitive to global features of the image for classification, simplifying the learning process and maintaining coherence in the annotation scheme.
By contrast, the annotation of object detection by YOLOv5 involves more fine-grained detail: every image includes one or more bounding boxes, with each being assigned a class of severity and accurate spatial coordinates. Dual-layer annotation thus enables the labeling of several lesions in a single image; this allows the model to provide both localization and classification, therefore offering clinically more relevant burn wound analysis..

Data Augmentation

A rich set of augmentation techniques is applied to the dataset, aiming to simulate real-world clinical variability that enhances model generalization and prevents overfitting. Geometric transformations including random rotations, different flips, and zoom operations provide variability in the spatial appearance of burns for handling different orientations and perspectives by models.
Photometric adjustments like contrast stretching and dynamic changes to HSV color channels introduce lighting and skin tone variabilities that simulate many different kinds of imaging environments patients may face in practice. Moreover, scaling techniques include random resizing and cropping to ensure the models are invariant to image size and focal regions. All these collectively prepare both CNN and YOLOv5 models for a wide range of clinical cases, thereby enhancing their utility and reliability in healthcare settings..
These augmentation strategies ensure sufficient variability to simulate diverse clinical conditions and imaging

characteristics, fostering generalization for deployment in practical healthcare scenarios.

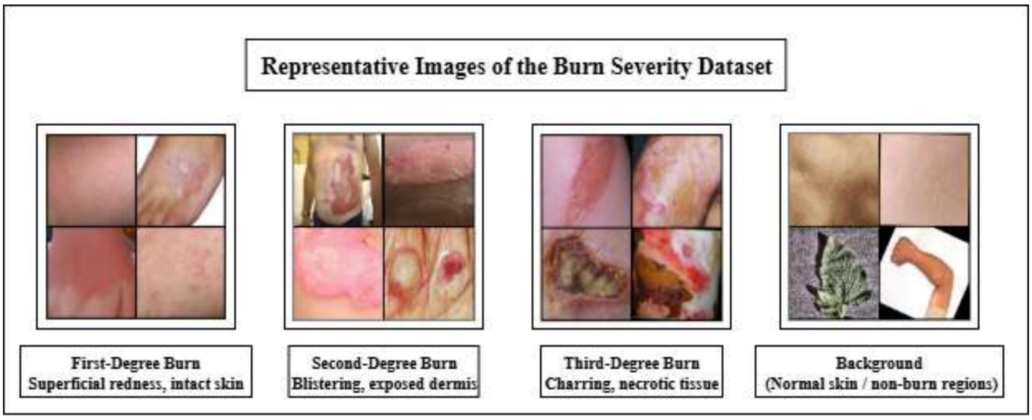

Fig 5: Representative images for each class in the dataset are depicted below

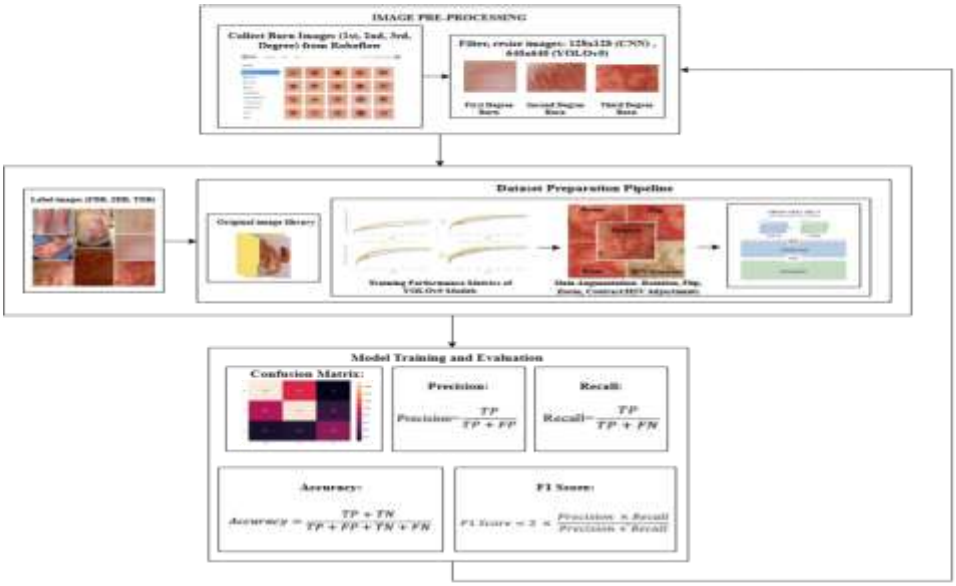

Fig 6: Flow Diagram

Results

Results summarized in Table 10 indicate a clear difference in the performance of shallow versus denser CNNs for burn classification. Comparing averaged metrics across accuracy, F1 score, recall, and precision, it is obvious that in all respects, the scores of the denser CNN are remarkably higher, while all the metrics have reached at least 0.75 or more. In contrast, the shallow CNN has always trailed behind with values around 0.65 to 0.66. This improvement signifies the benefits of increased network depth, which enables richer feature extraction and more reliable pattern recognition on medical imaging. Data indicate strong evidence for the application of deeper CNNs compared to shallower designs in complex tasks like automated analysis of burn severity, where subtle image variations can be critical in diagnosis and clinical decisions.

Table 10: Analytical aspects of shallow and deep CNN

Analytical Aspects	Shallow CNN	Denser CNN
Accuracy	66	76
F1 score	0.65	0.76
Recall	0.66	0.75
Precision	0.65	0.75

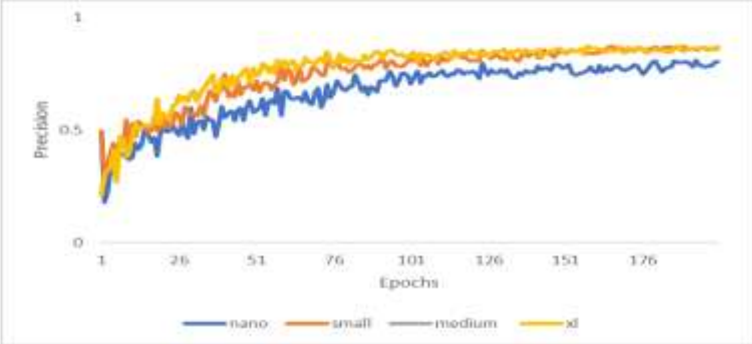

Fig. 7. Precision

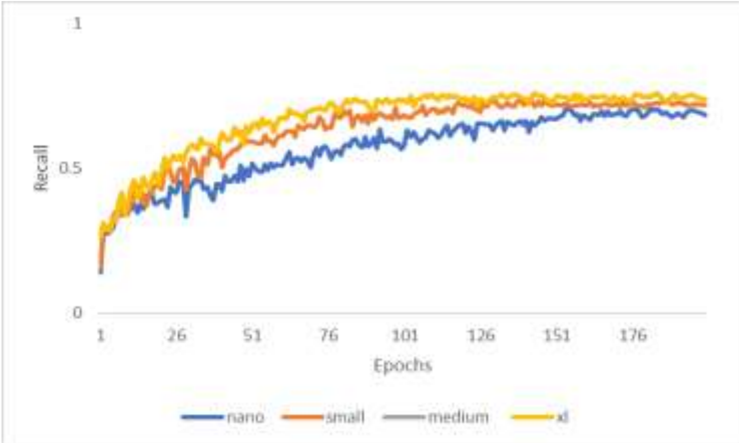

Fig. 8. Recall

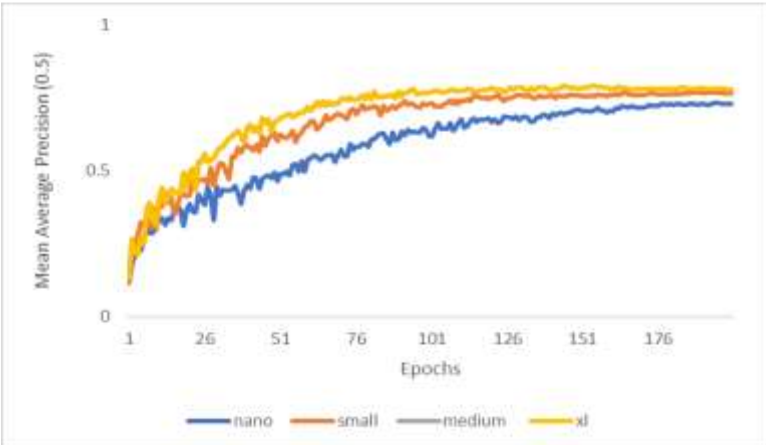

Fig.9.Mean Average Precision (0.5)

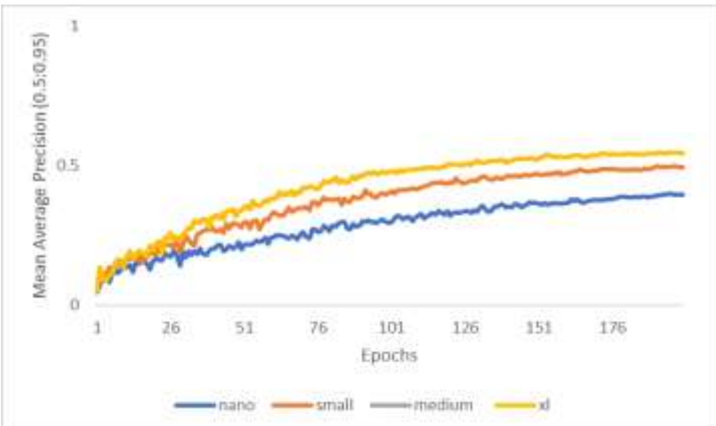

Fig. 10. Mean Average Precision (0.5:0.95)

Table 11: YOLOv5 Nano Predictive Scores

Nano	True 0	True 1	True 2	Background
Pred 0	0.77	0.01	0.02	0.41
Pred 1	0.02	0.78	0	0.15
Pred 2	0	0	0.7	0.44
Background	0.21	0.2	0.27	0

During model performance evaluation of the Nano YOLOv5 variant, one-vs-rest confusion matrices were created for all classes, and their respective Matthews Correlation Coefficient (MCC) values were calculated. The model was correct in classifying 0.77 samples as true positives for Class 0 while classifying 0.44 samples as false negatives, thus giving MCC as 0.5868. The classifier was most discriminant for Class 1 with 0.78 true positives while having only 0.17 as false negatives, giving the highest MCC as 0.7413. Class 2 obtained 0.70 true positives with 0.44 as false negatives, thus giving MCC as 0.5354. Generally, these values show that the Nano YOLOv5 variant has balanced predictive capacity in all classes with Class 1 being the most dependable, while Classes 0 and 2 are moderately consistent in classification power.

Table 12: YOLOv5 Small Predictive Scores

Small	True 0	True 1	True 2	Background
Pred 0	0.8	0.01	0.01	0.44
Pred 1	0.01	0.76	0	0.37
Pred 2	0.01	0	0.73	0.19
Background	0.18	0.23	0.26	0

During model performance evaluation of the Small YOLOv5 variant, one-vs-rest confusion matrices were created for each of the three classes, and their respective Matthews Correlation Coefficient (MCC) values were calculated. For Class 0, the model was correct in predicting 0.8 samples as true positives, while misclassifying 0.46 samples as false negatives, giving an MCC value of 0.6028. For Class 1, the classifier performed moderately, with 0.76 true positives and 0.38 false negatives, giving an MCC value of 0.6075. Lastly, Class 2 obtained the highest discriminative power, with 0.73 true positives and only 0.2 false negatives, giving the highest MCC value of 0.68. Generally, these findings suggest that the small YOLOv5 variant has a stable predictive performance in all classes, with Class 2 exhibiting the most dependable performance, while Classes 0 and 1 provide similar and robust classification strength.

Table 13: YOLOv5 Medium/XLPredictive Scores

Medium/XL	True 0	True 1	True 2	Background
Pred 0	0.79	0.01	0.01	0.43
Pred 1	0.01	0.76	0	0.34
Pred 2	0	0	0.74	0.23
Background	0.19	0.22	0.25	0

During model performance evaluation of the Medium & XL YOLOv5 models that both produced similar performance, one-vs-rest confusion matrices were calculated for all three classes, while their respective Matthews Correlation Coefficient (MCC) values were also estimated. For Class 0, the models accurately predicted 0.79 samples as true positives while having 0.45 samples as false negatives, giving an MCC value of 0.6044. For Class 1, the classifiers performed steadily, with 0.76 true positives as well as 0.35 false negatives, translating to an MCC value of 0.6272. Lastly, Class 2 recorded the best discriminative power with 0.74 true positives while having only 0.23 as false negatives, giving the highest MCC value of 0.6696. Generally, these outcomes suggest that both the medium and XL versions stay in good balance in their predictive power among all classes, with Class 2 exhibiting the best reliability, while Classes 0 & 1 have similar and good classification strength.

Discussion

This manuscript presents a detailed comparative analysis between the YOLOv5 deep learning approach and Convolutional Neural Networks (CNNs) for skin burn severity classification on a completely automatic basis. The results indicate that YOLOv5 as a single-stage object detector possesses a significant advantage over CNN architectures in terms of precision, accuracy, recall, and F1 score for all severities of burns. Most remarkably, YOLOv5's ability to perform multi-instance localization and classification of multiple burn lesions within a single image addresses a significant clinical void of conventional CNN classifiers, who provide a single holistic classification without a reference to a site.YO LOv5's improved performance can be attributed to its high-end architecture components including a 53-layer CSP backbone, Spatial Pyramid Pooling, and two prediction heads for bounding box regression and lesion-wise classification. These components allow for accurate definition of the boundaries of burn wounds, thus facilitating clinically meaningful measures such as estimations of Total Body Surface Area (TBSA) and lesion-based grade of burns necessary for treatment planning and follow-up.

While high-performing CNN architectures achieved moderate accuracy and are computationally less demanding than YOLOv5, low capacity to generalize to complex multi-lesion scenarios and real-time inference prevented them from being suitable for emergency settings. The study also reveals the adaptive value of efficient image augmentation methods, which improved model generalizability to different clinical scenarios.

Limitations are the reliance on a single annotated dataset and the computational capacity for training and deploying YOLOv5 models. Potential clinical verification of the latter in the future, enrichment of datasets with multi-center groups, and integration of automated classification output within clinical decision systems for burns care are regions for further study.

Conclusion

The current study illustrates the significant potential that YOLOv5 holds for automated burn grading capable of delivering precise and real-time lesion-based identification and classification of skin burns. Unlike conventional CNN-based methods, which offer single-label outputs with no spatial localization, the object detection architecture of YOLOv5 allows for the simultaneous classification and localization of multiple burn regions in a single image. Such comprehensive output closely aligns with the needs of clinical practice, given the importance of evaluating both the extent and severity of multiple lesions in patient management. Across experimental results and recent literature, YOLOv5 consistently showed superior performance, with higher accuracy, recall, and F1-scores in automated burn detection when compared to both shallow and deeper CNN approaches, thus ensuring higher discriminative capacity in diverse, real-world imaging scenarios.

From a practical clinical perspective, such deep learning models can significantly enhance the current standards of burn assessment, which still heavily rely on subjective visual inspections and clinician experience. Fast inference from YOLOv5, combined with balanced precision and recall across classes, presents an opportunity to bring practical value for medical triage, treatment planning, and monitoring. That includes additional benefits like reproducibility and scalability, both of which are important features for the standardization of care and reduction of the likelihood of misdiagnosis, with which provider fatigue, limited expertise, or inconsistent resource availability in health centers is often strongly associated. The ability of YOLOv5 to operate accurately in real time presents opportunities for embedding these tools in mobile health applications or point-of-care systems, thus extending specialized diagnostic capacity to remote or resource-limited settings.

This work underlines some key future directions in the field of automated burn diagnosis. Although YOLOv5 and other models have set high performance bars, open challenges include ensuring dataset diversity, model generalizability, and robust performances on edge cases-atypical burns, different skin tones, and confounding rare pathologies. Similarly, though promising accuracies are achieved in controlled datasets, real-world deployment would necessitate multi-center, multi-ethnic image repositories and prospective clinical trials. Specifically, model bias reduction, explainability, and transparency in algorithmic decision-making would also be important for trust and adoption by clinicians.

Looking ahead, the future scope for this research is promising on a number of fronts. This could be further enhanced by the integration of standard imaging with contextual clinical metadata, patient history, and even thermal or hyperspectral imaging to further improve diagnostic accuracy and patient-specific predictions. Development of interpretable AI, including attention maps or explainable feature attributions, may help bridge the gap between clinicians and algorithmic outputs, fostering greater confidence in AI-driven assessments. These should be supported by collaborative research involving medical practitioners, engineers, and data scientists that will be essential for refining architectures, optimizing user interfaces, and developing regulatory frameworks for clinical AI systems. Fundamentally, the iterative improvements of AI-powered diagnostic models such as YOLOv5 will serve to increase the reliability and accessibility of burn diagnostics, but it will also provide a framework for the integration of AI into emergency and critical care medicine more broadly. Adoption of these innovations has the potential not only to advance computational imaging but also to reconfigure pathways to timely, accurate, and equitable care in the management of acute skin injuries.

References

1. Abubakar, A., Ugail, H., Smith, K. M., Bukar, A. M., Elmahmudi, A. (2020). Burns depth assessment using deep learning features. Journal of Medical and Biological Engineering, 406, 923–933. https://doi.org/ 10.1007/s40846-020-00574-z
2. Precision Diagnosis of Burn Injuries Using Imaging and Predictive Modeling for Clinical Applications.0.1038/s41598-025-92096-4
3. Rahman, S., Rahman, M. M., Faezipour, M., Rastgaar, M., Ridelman, E., Klein, J. D., Angst, B. A., Shanti, C. M. (2024). Enhancing burn severity assessment with deep learning: A comparative analysis and computational efficiency evaluation. IEEE Access, 12, 147249–147262. https://doi.org/10.1109/ACCESS.2024.3476110
4. Deep Learning–Assisted Burn Wound Diagnosis: Diagnostic Model Development Study. 10.2196/22798
5. Moura, F. S. E., Amin, K., Ekwobi, C. (2021). Artificial intelligence in the management and treatment of burns: A systematic review. Burns & Trauma, 9, tkab022. https://doi.org/10.1093/burnst/tkab022
6. Jeschke, M. G., van Baar, M. E., Choudhry, M. A., Chung, K. K., Gibran, N. S., Logsetty, S. Burn injury. https://doi.org/10.1038/s41572-020-0145-5
7. Siu, W. S., Ma, H., Leung, P. C. Review on current advancements in facilitation of burn wound healing. DOI:10.3390/bioengineering12040428
8. G. Jocher et al., YOLOv5 by Ultralytics, 2021. Online. Available: https://github.com/ultralytics/yolov5
9. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. doi = {10.1145/3065386}
10. Redmon, J., Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.
11. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).DOI: 10.1109/CVPR.2016.90
12. Tan, M., Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (pp. 6105–6114).DOI: 10.48550/arXiv.1905.11946
13. Sharma, J., Kumar, D., Bhatnagar, S. (2024). A YOLO-based object detection model for counting the number of mathematical expressions and symbols. In 2nd International Conference, 23 October 2024.DOI: 10.1109/ICSSAS64001.2024.10760799
14. Usman, R. M., Yan, J., Jamil, M. (2024). Enhancing object detection and classification using deep learning: A novel approach with annotated dataset and YOLO-C3GTR architecture. Mehran University Research Journal, April 2024.DOI: 10.22581/muet1982.3186
15. Xu, K., Sun, W., Yang, R. (2024). Early Sweet Potato Plant Detection Method Based on YOLOv8s ESPPD-YOLO: A Model for Early Sweet Potato Plant Detection in a Complex Field Environment. Agronomy, 11 November 2024. DOI: 10.3390/agronomy14112650
16. Carrasco, D. P., Rashwan, H. A., Puig, D. (2021). T-YOLO: Tiny vehicle detection based on YOLO and multi-scale convolutional neural networks. IEEE Access. 10.1109/ACCESS.2021.3137638
17. Sharma, P., Jain, G., Joshi, S. (2024). Multi-class classification and detection of brain tumor ROI: Yolov9–CNN. International Conference, 23 August 2024.10.1109/SPICSCON64195.2024.10941675
18. Sakib, S., Ahmed, N., Ahmed, H. (2018). An overview of convolutional neural network: Its architecture and applications. Published November 2018. 10.20944/preprints201811.0546.v1
19. Singla, M., Gill, K. S., Rawat, R. S. (2024). Transfer learning-based lung opacity classification in chest X-ray image dataset utilizing VGG16 pre-trained CNN model. Asia Pacific Conference on Innovation in Technology, July 2024.10.1109/APCIT62007.2024.10673688
20. Bhuria, R., Gill, K. S., Dangi, S. (2024). Utilizing a pre-trained VGG16 CNN model: Transfer learning-based lung opacity classification in chest X-ray dataset. IEEE International Conference, July 2024.10.1109/APCIT62007.2024.10673688
21. Naman, S., Sharma, S., Baldi, A. (2024). A CNN-based machine learning model for the identification of herbal drugs: A case study on cumin. Current Computer Science, July 2024. 10.2174/0129503779301717240607070206
22. Priya, P., Vaishnavi, T., Reethika, A. (2023). An enhanced animal species classification and prediction engine using CNN. 2nd International Conference on Edge Computing and Applications. DOI: 10.1109/ICECAA58104.2023.10212299