ISSN : 2455-9679
SJIF Impact Factor : 6.190



IJRTSM
INTERNATIONAL JOURNAL OF RECENT TECHNOLOGY SCIENCE & MANAGEMENT

"HYBRID DEEP LEARNING-BASED DOG SKIN CONDITION CLASSIFICATION USING IMAGES"

Deepak Thakre [1], Dr. Dilip Solanki [2]
[1] P. G. Scholar, Institute of Advance Computing, SAGE University, Indore, Madhya Pradesh, India
[2] Professor, Institute of Advance Computing, SAGE University Indore, Madhya Pradesh, India

ABSTRACT

Canine skin disorders can present with overlapping visible signs, including erythema, alopecia, scaling, crusting, papules, and changes in skin texture. Image-based classification may help organize visual observations and support preliminary triage, but it cannot replace veterinary examination, laboratory testing, or clinical judgement. This paper describes a hybrid deep-learning pipeline designed to classify dog skin-condition images into four project-defined categories: bacterial dermatosis, fungal infections, healthy skin, and hypersensitivity/allergic dermatosis. The supplied dataset analysis identifies 442 readable images distributed across these categories. The project combines EfficientNetB0 and MobileNetV2 feature extractors, pools their feature maps, concatenates the resulting vectors, and passes the fused representation through batch normalization, dropout, and a four-unit softmax output layer. The dataset workflow includes image-manifest generation, label-review flags, near-duplicate analysis, and a group-aware stratified train/validation/test split intended to reduce leakage from related or augmented images. The planned training procedure uses transfer learning when pretrained weights are available, data augmentation, class weighting, early stopping, learning-rate reduction, and checkpointing. Evaluation is intended to include accuracy, class-wise precision and recall, macro and weighted F1 scores, confusion matrices, and visual explanation using Grad-CAM. However, the supplied project status states that training and evaluation have not yet been run. Consequently, this manuscript reports the dataset and planned method rather than validated predictive performance. The work emphasizes reproducibility, transparent limitations, and the need for external clinical validation before any practical veterinary use.

Key Words: *Canine dermatology; skin-condition classification; deep learning; EfficientNetB0; MobileNetV2; transfer learning; image classification; Grad-CAM.*

I. INTRODUCTION

Skin and coat abnormalities are common reasons for veterinary consultation. Their visible appearance can be influenced by infection, allergy, parasites, trauma, environmental exposure, grooming, breed characteristics, and secondary inflammation. Different causes may produce similar signs, while a single underlying condition may appear differently depending on disease stage, coat colour, lighting, anatomical site, and prior treatment. Visual assessment is therefore useful but often insufficient to establish a definitive diagnosis.

Digital image analysis offers a way to organize visual information consistently and to investigate whether patterns in photographs can be learned by machine-learning systems. Convolutional neural networks (CNNs) are particularly suitable for image classification because they learn hierarchical representations of colour, texture, edges, and higher-

ISSN : 2455-9679
SJIF Impact Factor : 6.190

level visual structures from training data. Transfer learning can make CNN-based development more practical when a task-specific dataset is small, because a network pretrained on a large image collection can provide reusable features.

This project investigates a two-backbone feature-fusion approach for four broad dog skin-condition labels. EfficientNetB0 is designed to balance model accuracy and computational efficiency through compound scaling, while MobileNetV2 uses inverted residual blocks and linear bottlenecks to provide an efficient representation. The project combines features from both networks rather than relying on a single backbone. The intention is to test whether complementary feature representations can support discrimination among the project's four categories.

The work must be interpreted as a computer-vision classification study, not as a diagnostic system that identifies a confirmed disease from a photograph. The categories are broad dataset labels and may encompass clinically heterogeneous presentations. In particular, the image label may not be based on a uniform diagnostic reference standard, and photographs may not capture information needed to distinguish bacterial, fungal, and allergic disease. Any eventual model output should therefore be described as a prediction of a dataset category rather than a veterinary diagnosis.

1.1 Problem statement

The main problem is to build a reproducible image-classification pipeline for a small and potentially heterogeneous canine skin-image dataset, while limiting overfitting and data leakage. The study must also report evaluation transparently. A high score on images that are duplicated, augmented from the same original photograph, or collected under similar conditions may overstate generalization. Dataset quality, partitioning strategy, and independent testing are therefore central parts of the method rather than secondary implementation details.

1.2 Aim and objectives

- Develop a hybrid image-classification architecture using EfficientNetB0 and MobileNetV2 feature extractors.
- Describe and audit the supplied dataset across four project-defined categories.
- Use group-aware partitioning to keep related images within the same data partition where source-group information is available.
- Apply transfer learning, augmentation, class weighting, and regularization to reduce overfitting.
- Define a transparent evaluation plan that includes class-wise metrics and error analysis.
- Investigate Grad-CAM visualizations as a qualitative check of image regions influencing predictions.
- State the limitations and evidence required before the method could be considered for clinical decision support.

1.3 Research questions

1. Can a dual-backbone feature-fusion model be implemented to predict the four labels represented in the supplied dataset?
2. How should the dataset be partitioned and audited to reduce the risk of leakage from duplicate or augmented images?
3. Which class-wise evaluation measures and visual checks are necessary to assess the model beyond overall accuracy?
4. What evidence is still required to determine whether predictions generalize to new dogs, clinics, cameras, and image conditions?

II. Related Work and Technical Background

2.1 Convolutional neural networks for image classification

CNNs learn spatially local patterns through convolutional filters and progressively combine these patterns into higher-level features. Early layers often respond to simple edges and colour transitions, while deeper layers can represent more complex textures and object parts. Pooling or global-average-pooling layers reduce spatial dimensions and create compact representations that can be passed to a classifier. For skin photographs, such features may capture lesion texture, redness, scaling, hair loss, or broader visual context. Yet these patterns are not necessarily disease-specific: camera flash, fur, background, image compression, and skin pigmentation can also influence a model.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

A dataset-specific CNN can be trained from random initialization, but doing so generally requires enough diverse labelled examples to learn stable features. When only a few hundred images are available, transfer learning is a common alternative. A model pretrained on a large generic image dataset can be adapted to a new task by replacing its classification head and training selected layers. This can reduce training requirements, although it does not guarantee good performance when the source and target image domains differ.

2.2 EfficientNetB0

EfficientNet proposes a family of convolutional networks that jointly scale network depth, width, and input resolution. EfficientNetB0 is the baseline member of the family and is often used when a balance between representational capacity and computation is desired. In the supplied implementation, EfficientNetB0 receives 224 × 224 RGB images and its convolutional output is passed through global average pooling to form a 1,280-dimensional feature vector. The code uses the model's built-in rescaling behaviour for this branch. When pretrained ImageNet weights are successfully loaded, the backbone can begin with general visual features; the project also includes a fallback path if the weights cannot be downloaded.

2.3 MobileNetV2

MobileNetV2 was designed for efficient inference and uses depthwise separable convolutions, inverted residual connections, and linear bottlenecks. These design choices reduce computational cost relative to many larger convolutional models. In the supplied code, MobileNetV2 also receives 224 × 224 RGB images and produces a 1,280-dimensional pooled feature vector. The input is rescaled to the range expected by the MobileNetV2 preprocessing pipeline. This branch provides a second learned representation that may respond differently to image structures than EfficientNetB0.

2.4 Feature fusion and regularization

Feature-level fusion combines representations from multiple feature extractors before the final classifier. Here, the two 1,280-dimensional vectors are concatenated to form a 2,560-dimensional representation. Batch normalization is then applied, followed by dropout with a rate of 0.4 and a dense softmax layer with four outputs. Concatenation allows the classifier to use information from both backbones, but it also increases the dimensionality of the classifier input and may overfit when training data are limited. Fusion should therefore be assessed empirically against single-backbone baselines rather than assumed to improve performance.

2.5 Explainability and Grad-CAM

Class Activation Mapping methods aim to highlight image regions associated with a model's class score. Grad-CAM uses gradients flowing into a convolutional layer to create a coarse localization map. Such heatmaps can help reviewers inspect whether a model appears to focus on a lesion or instead relies on background, watermarks, image borders, or other shortcuts. A heatmap is not proof that a model's reasoning is clinically correct, nor does it establish causality. It should be treated as a qualitative diagnostic aid alongside error analysis.

2.6 Position of the present project

The contribution described here is the design and audit of a hybrid EfficientNetB0–MobileNetV2 pipeline for a small four-category dog skin-image dataset, with explicit attention to grouped splitting and transparent evaluation. The supplied project files do not yet contain trained model weights or executed test results. Therefore, the present manuscript does not claim that the hybrid architecture outperforms a single model, nor does it claim a particular level of accuracy. These claims can only be made after the planned experiments are completed and documented.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

III. MATERIALS AND METHODS

3.1 Study design

This is a supervised, image-based, four-class classification project. Each image is associated with one project label, and the model is trained to estimate a probability distribution over those labels. The methodological workflow consists of dataset inspection, duplicate and group analysis, partition generation, image preprocessing, model construction, training, test-set evaluation, and qualitative explanation. The steps below describe the supplied implementation and intended evaluation plan; they should not be interpreted as evidence that training has already completed.

3.2 Dataset and class composition

The dataset analysis report supplied with the project identifies 442 readable images. The four labels and reported counts are shown in Table 1. These counts sum to 442. The image collection contains JPG and WebP files, with most images stored as JPG. Image dimensions vary, so resizing is required before batch training. The dataset's size is modest for deep learning and the classes are not perfectly balanced.

Class label in project	Number of readable images	Share of dataset
Bacterial_dermatosis	97	21.9%
Fungal_infections	137	31.0%
Healthy	119	26.9%
Hypersensitivity_allergic_dermatosis	89	20.1%
Total	442	100.0%

The class names are reproduced from the project rather than reinterpreted as standardized clinical diagnoses. The paper does not have access to patient records or a uniform diagnostic protocol confirming the labels. Consequently, label correctness and clinical class definitions remain limitations that should be addressed before stronger conclusions are drawn.

3.3 Dataset quality and duplicate analysis

The supplied dataset-analysis report states that no exact duplicates were detected using MD5 file hashes. It also reports 10 near-duplicate pairs, including one pair whose images appear assigned to different classes. The report identifies 126 files associated with Roboflow-augmented copies of 42 source-image groups and flags nine files for potential label review. These flags are audit indicators, not definitive proof that a label is wrong. The flagged images should be reviewed by a qualified person, and ambiguous cases should be resolved using the best available source information before final training.

Exact hashing detects byte-identical files but will not detect a resized, cropped, compressed, colour-adjusted, or otherwise transformed version of the same photograph. Perceptual similarity and known augmentation lineage are therefore important. If two related images are split across training and test sets, the test score may be artificially high because the model has effectively seen nearly the same visual content during training. The project's group-aware split is intended to keep known related images together.

3.4 Data partitioning

The project records a split of 315 training images, 64 validation images, and 63 test images, corresponding approximately to 71.3%, 14.5%, and 14.3% of the 442-image dataset. The split-generation configuration uses a fixed random seed of 42 and a group-aware stratified strategy. Group awareness aims to prevent related or augmented members of a source-image group from crossing the partition boundaries, while stratification aims to preserve class representation as much as possible.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

Partition	Images	Approximate proportion	Purpose
Training	315	71.3%	Parameter learning and augmentation
Validation	64	14.5%	Model selection, early stopping, learning-rate scheduling
Test	63	14.3%	Final evaluation after decisions are frozen
Total	442	100.0%	Dataset audit total

The split metadata indicates that no recorded groups cross partitions. However, the project does not contain dog identifiers, so dog-level independence cannot be established. Images of the same dog that are not linked by source-group metadata could still occur in different partitions. The test set should therefore be described as a held-out image-level test set with group-aware controls, not as a confirmed independent cohort of unseen dogs.

The test set must not be used to choose hyperparameters, adjust thresholds, select a checkpoint, or repeatedly revise the model. If repeated experimentation uses test performance to guide decisions, the test set becomes part of the development process and its reported score loses its status as an unbiased final estimate. For stronger evidence, a later study should use an external dataset collected from different dogs, clinics, and acquisition conditions.

3.5 Image preprocessing

Images are decoded as RGB and resized to 224 × 224 pixels for the model input. Resizing produces a consistent tensor shape but can distort aspect ratio unless a crop or padding strategy is used; the exact policy should be kept consistent and documented. The EfficientNetB0 branch uses its built-in rescaling behaviour, while the MobileNetV2 branch rescales input values to the range expected by that backbone. Separate preprocessing paths are important because pretrained networks may expect different pixel-value ranges. Any future implementation change should verify that preprocessing matches the selected pretrained weights.

Data augmentation is applied during training to expose the model to plausible image variation and reduce reliance on incidental features. Augmentation should be limited to transformations that preserve the class meaning. For example, modest geometric transformations may be acceptable, whereas aggressive colour shifts could alter redness or skin tone cues relevant to the task. Augmentation is intended only for training data; validation and test images should be evaluated without random augmentation, apart from deterministic resizing and normalization.

3.6 Proposed model architecture

The hybrid model contains two parallel convolutional backbones. EfficientNetB0 and MobileNetV2 each process the same resized RGB image. Their final convolutional representations are globally average pooled, resulting in one 1,280-dimensional vector per backbone. The vectors are concatenated into a 2,560-dimensional feature vector. Batch normalization stabilizes the fused activations, dropout with rate 0.4 provides regularization, and a dense layer with four softmax units generates class probabilities.

Component	Output / setting	Role
Input	224 × 224 × 3 RGB	Standardized image tensor
EfficientNetB0 backbone	Feature extractor	General and fine-grained visual representation
Global average pooling	1,280 features	Compress spatial feature maps
MobileNetV2 backbone	Feature extractor	Second, computationally efficient representation
Global average pooling	1,280 features	Compress spatial feature maps
Concatenation	2,560 features	Fuse both backbone representations
Batch normalization	2,560 normalized activations	Stabilize fused representation
Dropout	Rate = 0.4	Reduce over-reliance on individual activations
Dense + softmax	4 probabilities	Predict the four dataset labels

ISSN : 2455-9679
SJIF Impact Factor : 6.190

This is feature concatenation, not an attention-based fusion module, transformer, object detector, or lesion segmentation network. The model predicts one class for the full image and does not explicitly localize the lesion. If the image contains multiple skin findings, or if the relevant lesion occupies only a small region, global classification may be influenced by background and coat features.

3.7 Transfer learning and optimization

The training configuration specifies a batch size of 16 and up to 15 epochs for the initial training stage, using a learning rate of 0.001. A fine-tuning stage can run for up to 15 additional epochs with a learning rate of 0.00001. The code attempts to load ImageNet-pretrained weights and includes a fallback if those weights cannot be obtained. The actual initialization used in any experiment must be recorded, because training from random initialization is not equivalent to transfer learning and may produce substantially different outcomes.

The configured optimizer is Adam and the loss function is sparse categorical cross-entropy. Class weights are used to reduce the influence of class imbalance on the loss. Early stopping and learning-rate reduction callbacks are included to help manage overfitting and stalled validation performance. The fine-tuning procedure is designed to unfreeze the last 30 layers of each backbone, subject to the implementation's layer-freezing rules. The final manuscript should report the actual epoch at which training stopped, the best validation checkpoint, and whether each backbone was initialized with pretrained weights.

A code-level reproducibility concern must be resolved before reporting results: the checkpoint callback is configured to save according to validation accuracy, while early stopping restores weights according to validation loss. These criteria can select different epochs. The final evaluation should use one clearly defined checkpoint-selection policy, and the exact saved model should be the model evaluated on the test set. Training logs should be retained to confirm this.

3.8 Evaluation metrics

Overall accuracy is the proportion of correctly classified test images. It is easy to interpret but can conceal poor performance in a smaller class. Precision for a class measures the proportion of images predicted as that class that truly belong to it under the dataset labels; recall measures the proportion of actual class images correctly identified. The F1 score is the harmonic mean of precision and recall. Macro-averaged scores give each class equal weight, while weighted averages account for class support. Because the dataset is modest and somewhat imbalanced, macro precision, macro recall, macro F1, and per-class confusion patterns should be emphasized alongside accuracy.

Metric	Definition / interpretation	Why it is reported
Accuracy	Correct predictions divided by all test predictions	Overall classification performance
Per-class precision	TP / (TP + FP) for each class	Shows reliability of class predictions
Per-class recall	TP / (TP + FN) for each class	Shows how many labelled examples of each class are recovered
Per-class F1	Harmonic mean of precision and recall	Balances false positives and false negatives
Macro F1	Unweighted mean of class F1 scores	Gives equal importance to all four labels
Weighted F1	Class-support-weighted mean of class F1 scores	Summarizes performance with class frequencies considered
Confusion matrix	Counts true labels against predicted labels	Reveals which categories are confused

The evaluation script also plans ROC and precision-recall curves. For a multi-class problem, these require an explicit one-vs-rest or micro/macro aggregation definition and probability outputs. Such curves can be informative, but with only 63 test images, estimates may be unstable and should be interpreted cautiously. Confidence intervals or repeated evaluation on larger independent datasets would improve the reliability of comparisons.

3.9 Explainability analysis

Grad-CAM is included as a qualitative inspection tool. Heatmaps should be generated for correctly and incorrectly classified examples, with examples selected across all four classes rather than only visually convincing cases.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

Reviewers should check whether salient regions overlap the visible skin finding and whether the model instead emphasizes background, fur, a hand, a collar, image borders, text, or other confounding content. Heatmaps should not be presented as evidence of clinical validity; they only show a gradient-based view of model sensitivity for a selected class and layer.

3.10 Reproducibility and implementation

The project uses Python and TensorFlow/Keras components, with supporting libraries for image handling, numerical operations, metrics, plotting, and the optional Streamlit prediction interface. A fixed random seed is specified, but exact reproducibility can still vary across library versions, hardware, and nondeterministic operations. The final study record should include the operating system, Python and package versions, GPU/CPU details, model initialization status, full configuration file, random seed, training logs, saved checkpoint hash, and test-evaluation output.

Before a result is considered final, the project should be run from a clean environment using the documented setup instructions. Unit tests should verify class-index mapping, preprocessing, output dimensions, prediction decoding, and the consistency of the saved model with the evaluation script. The status file supplied with the project states that the code had been syntax-checked but that dependencies and full runtime execution had not been successfully validated in the original environment. A successful syntax check is not equivalent to a successful training run.

IV. RESULTS

4.1 Dataset audit results

The dataset audit supports a descriptive summary of the supplied image collection. There are 442 readable images across four labels, with the fungal-infection category containing the largest share (137 images; approximately 31.0%) and the hypersensitivity/allergic-dermatosis category containing the smallest share (89 images; approximately 20.1%). The imbalance is moderate rather than extreme, but it justifies reporting class-wise performance and using the configured class-weighting strategy as an experimental choice.

The audit also indicates that related images require careful handling. The report records 10 near-duplicate pairs, one cross-class near-duplicate pair, and nine files flagged for possible label review. In addition, 126 images are associated with augmented copies of 42 source-image groups. These findings support the decision to use group-aware partitioning and to inspect suspicious pairs before drawing conclusions. They do not, by themselves, establish that all labels are correct or that the dataset represents independent dogs.

Audit item	Reported finding	Interpretation
Readable images	442	Dataset size used for the reported split
Exact duplicates	None detected by MD5	Byte-identical duplicates not found
Near-duplicate pairs	10	Visual similarity requires review
Cross-class near-duplicate pair	1	Potential label or grouping inconsistency
Potential label-review flags	9 files	Manual verification recommended
Augmentation/source grouping	126 files associated with 42 source groups	Related images should remain in one partition
Dog-level identifiers	Not available	Independent-dog generalization cannot be verified

4.2 Model performance status

No numerical model-performance results are reported in this manuscript because the supplied project status explicitly states that training and evaluation have not yet been executed. No trained model files, completed test metrics, confusion matrix, ROC/PR plots, learning curves, or Grad-CAM outputs were available in the supplied project package. It would be scientifically inappropriate to insert estimated or illustrative scores as if they were observed results.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

V. DISCUSSION

5.1 Interpretation of the proposed approach

The two-backbone design offers a reasonable experimental strategy for exploring complementary visual representations. EfficientNetB0 and MobileNetV2 were selected because they provide established convolutional feature extractors with different architectural priorities. Concatenating their pooled representations gives the final classifier access to both feature vectors. However, the design also increases the number of parameters in the classification head and may increase inference cost. Its value must be established by comparison, not assumed from architectural complexity.

A suitable experiment would compare the hybrid model with EfficientNetB0 alone and MobileNetV2 alone under the same data partitions, preprocessing, and checkpoint policy. If the hybrid model achieves a higher score, the analysis should determine whether the gain is consistent across classes and whether it justifies additional memory and latency. If the hybrid model does not improve results, the simpler backbone may be preferable. Fair comparison requires using the same held-out test set only after all design decisions have been finalized.

5.2 Importance of dataset integrity

For a small image dataset, the way images are grouped and split can influence reported performance as much as the choice of architecture. Augmented versions of a source image share lesion geometry, fur arrangement, lighting, and background. If such images cross partitions, the test set may not represent genuinely new visual evidence. The project's group-aware split is a useful safeguard, but it only protects against relationships that have been identified. Missing dog IDs and unknown collection provenance remain important sources of uncertainty.

The cross-class near-duplicate pair and potential label-review flags deserve special attention. A near-duplicate pair assigned to different classes can arise from an annotation error, a genuinely ambiguous example, a different interpretation of the image, or an imperfect similarity flag. Each case should be reviewed rather than automatically deleted or relabelled. A transparent audit log should record the image identifiers, reason for review, reviewer decision, and whether the decision was made before test-set evaluation.

5.3 Clinical meaning of the labels

The four labels represent broad categories, and the image alone may not contain enough information to distinguish them reliably. Bacterial and fungal conditions may share redness, crusting, hair loss, or scaling; allergic skin disease may be complicated by secondary infection. A healthy-versus-abnormal distinction may also be affected by image quality, body site, and coat characteristics. A model can learn correlations in the dataset without learning a clinically valid distinction. Accordingly, predictions should be framed as category estimates for research purposes and not as a replacement for veterinary diagnosis.

A stronger future dataset would link each image to a documented diagnostic reference standard, such as a veterinarian's assessment supported where appropriate by cytology, culture, skin scraping, allergy work-up, or other tests. The precise reference standard will differ by suspected condition. Annotation guidelines should specify whether each image receives a single label, multiple labels, or an uncertainty designation. Inter-rater agreement and adjudication procedures should also be reported.

5.4 Generalization and external validation

A held-out subset from the same source distribution measures performance on images that resemble the training collection. It does not establish performance on new clinics, new cameras, different lighting, different breeds, varying coat colours, or different disease severity. External validation should therefore be performed on images collected independently and ideally from multiple veterinary settings. The external test should be locked before model selection and should include enough cases in each class to estimate class-wise metrics with useful precision.

The intended use must also be specified. A tool for educational exploration has a different risk profile from a triage tool used by owners or veterinary staff. Before any decision-support use, a prospective evaluation should examine false-negative and false-positive consequences, calibration of predicted probabilities, subgroup performance, usability, and whether access to model predictions changes clinical decisions. Human factors and escalation rules are as important as offline accuracy.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

5.5 Explainability and failure analysis

Grad-CAM can reveal suspicious attention patterns, but it should be paired with direct review of errors. False negatives may be particularly important if a disease class is overlooked; false positives may lead to unnecessary concern or inappropriate self-treatment. For each error category, reviewers should examine image focus, exposure, lesion size, anatomical site, visible background, class ambiguity, and label confidence. Results should include examples of failures as well as successes, with identifying or sensitive metadata removed.

5.6 Efficiency and deployment

MobileNetV2 is often used in efficiency-constrained settings, while the hybrid model requires both backbones to run for each image. Whether the combined model is suitable for a desktop or mobile deployment depends on measured model size, memory use, inference latency, and device hardware. The supplied project includes a Streamlit interface for demonstration, but an interface does not itself establish clinical readiness. Any deployment should clearly display that the output is an unvalidated model prediction and should direct users to seek veterinary advice for concerning symptoms.

5.7 Limitations

- The dataset contains only 442 readable images, limiting statistical power and diversity.
- The supplied project does not include dog-level identifiers, so separation by individual animal cannot be confirmed.
- Labels are reproduced from the dataset; a uniform clinical reference standard is not documented in the supplied files.
- Near-duplicate pairs and potential label-review cases require manual resolution.
- Training and final evaluation have not yet been executed, so no performance claims can be made.
- The pretrained-weight download may fail and trigger a random-initialization fallback; the actual initialization must be recorded.
- Checkpoint saving and early-stopping restoration use different validation criteria in the supplied code and should be reconciled.
- Grad-CAM is qualitative and cannot establish clinical reasoning or diagnostic validity.
- The project has not been externally validated on images from independent dogs, clinics, or acquisition settings.

VI. Recommended Experimental Protocol Before Submission

- The following checklist should be completed before this document is submitted as a completed empirical research article. It is included to prevent the methodology description from being mistaken for executed results.
- Review the nine flagged files and the cross-class near-duplicate pair; record decisions and keep an audit trail.
- Verify that source groups and augmented copies do not cross train, validation, and test partitions.
- Confirm the class-index mapping used by training, evaluation, and prediction scripts.
- Run the pipeline in a clean environment and record Python, TensorFlow/Keras, and dependency versions.
- Confirm whether ImageNet weights loaded successfully for both backbones; record any fallback to random initialization.
- Align checkpoint saving and early stopping to one clearly stated selection criterion.
- Train the model and save logs, configuration, model checkpoint, and random seed.
- Evaluate the chosen checkpoint once on the locked test set and export all metrics and plots.
- Compare the hybrid model against each backbone individually under a fair, identical protocol.
- Generate Grad-CAM examples across all classes and inspect both correct and incorrect predictions.
- Report model size and inference latency on clearly specified hardware.
- Add the actual results to Section 4 and revise the abstract, discussion, and conclusion to match the evidence.
- If possible, conduct external validation on a dataset from independent dogs and sources before making claims about real-world generalization.

ISSN : 2455-9679
SJIF Impact Factor : 6.190

XI. CONCLUSION

This manuscript documents a hybrid deep-learning pipeline for classifying dog skin-condition images into bacterial dermatosis, fungal infections, healthy skin, and hypersensitivity/allergic dermatosis. The supplied dataset audit identifies 442 readable images and highlights class imbalance, near-duplicate pairs, potential label-review cases, and the need to keep augmented source groups within a single partition. The model architecture combines 1,280-dimensional pooled features from EfficientNetB0 and MobileNetV2 into a 2,560-dimensional representation, followed by batch normalization, dropout, and a four-class softmax classifier. The planned training process uses transfer learning where available, augmentation, class weighting, and validation-based controls.

The principal conclusion at this stage is methodological rather than performance-based. The supplied project files do not contain executed training or test results, so no accuracy or superiority claim is justified. Before publication, the model must be run, the checkpoint-selection inconsistency must be resolved, the flagged data must be reviewed, and all test metrics must be reported transparently. External validation and clinically grounded labels are required before the approach can be considered for veterinary decision support.

REFERENCES

[1] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research, 97, 6105–6114.

[2] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4510–4520.

[3] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626.

[4] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248–255.

[5] 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, 770–778.