Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 62(1): 53-65 (2025)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(62-1)681

Pakistan Academy of Sciences

Research Article

Evaluating the Efficacy of Convolutional Neural Networks Across Diverse Datasets

Swati Gupta*, Bal Kishan, and Pooja Mittal

Department of Computer Science and Applications, Maharshi Dayanand University (MDU), Rohtak, India

Abstract: Focusing on sentiment analysis and medical image processing, this paper assesses Convolutional Neural Networks performance across different datasets. Emphasized are recent developments in deep learning models for segmentation and picture categorization as well as other purposes. Using contemporary Convolutional Neural Network architectures, this work seeks to get good performance in medical diagnosis and sentiment analysis. The paper emphasizes how well Convolutional Neural Networks perform in domain-specific tasks. Using the same data preparation techniques, appropriate designs, and strategies to split the data, this paper investigates how well Convolutional Neural Network algorithms perform on medical data and sentiment analysis. Convolutional Neural Network models are optimized using hyperparameter tweaking and cross-validation techniques. While guaranteeing patient privacy, data anonymization, and bias reduction, the research seeks to highlight strengths, weaknesses, and patterns. Focusing on ethical concerns and offering suggestions for improvement, it tackles problems in sentiment categorization and medical imaging anomaly detection. This work attains 96 percent accuracy using Convolutional Neural Networks across four datasets. Common measures in the performance assessments of Sentiment Analysis, Skin Cancer Detection, Brain Tumour Detection, and Kidney Stone Detection include F1 scores, recall, and accuracy. With 0.97, Brain Tumour Detection had the highest accuracy; Kidney Stone Detection and Skin Cancer Detection both had 0.95; Sentiment Analysis scored 0.96. The consistently high recall and accuracy scores across all domains indicate good classification capabilities; an F1 score between 0.95 and 0.96 guarantees outstanding performance in both detection and analysis tasks.

Keywords: Accuracy, Precision, Performance Metrics, Diverse Datasets, CNNs.

1. INTRODUCTION

CNNs are special neural networks that can process inputs that resemble grids, such as images. CNNs leverage a basic design of connected layers to enable tasks like object recognition and picture categorisation. These activities rely on the input data to extract key attributes. CNNs excel in part because they can learn feature hierarchies from input pictures automatically and efficiently. Convolution layers do this by applying filters to the input picture to identify different patterns and characteristics [1]. To further reduce computational complexity while preserving critical information. In the end, the learned representations in Figure 1 are used to make predictions, and the retrieved features are integrated by fully linked layers. The figure shows CNN architecture for image processing or recognition tasks. It consists of an input, convolutional, max-pooling, hidden, and an output layer. Compared with 125 x 125 pixels, the input layer takes color pictures. Max-pooling layers downscale feature maps; convolutional layers shrink spatial dimensions. Compared with 784 and 16 neurons, the hidden layers enable the learning of intricate patterns. The S neurons in output layer reflect classes in the classification challenge [1].

Although very effective, CNNs have several shortcomings. One disadvantage of vanishing gradients in deep networks with several layers is that the gradients required changing network parameters decrease, thereby either delaying or ceasing learning. Usually, proper network weight

Received: November 2024; Revised: February 2025; Accepted: March 2025
* Corresponding Author: Swati Gupta <swati.rs20.dcsa@mdurohtak.ac.in>

initialization, specialized activation functions, batch normalization, and skip connections help to encourage gradient flow throughout the network. Current deep learning methods depend on CNNs, which also drive image analysis and related fields to unprecedented heights despite challenges. CNNs shine in NLP and computer vision. CNNs are the foundation of healthcare image analysis in disease detection and identification [1]. CNNs can assist in the image data analysis for public opinion and sentiment trends. This work will evaluate the performance of the CNN algorithm in skin cancer detection [2], emotion analysis [3], renal stone detection [4], and brain tumor diagnosis [5]. The joint study of these domains clarifies CNN efficiency in various situations and shares obstacles and discoveries are shown in Figure 1.

Figure 1 presents Convolutional Neural Network (CNN) architecture where input layer receives raw image data, which is passed through convolutional layers that apply filters to extract local features like edges and textures. These features are then passed through activation functions such as ReLU to introduce non-linearity, followed by pooling layers that reduce the spatial size and retain important information. This feature extraction process is repeated to form a deep representation of the input. The resultant feature maps are flattened and transmitted to fully linked layers, which identify high-level patterns and connections. The output layer uses methods such as softmax to categorise the input into established classifications. The significance of the study lies in the comprehensive assessment of CNNs in two crucial areas: sentiment analysis [3] and medical image processing [4, 5]. This work improves diagnostic accuracy, fraud prevention, and sentiment analysis by addressing the limitations of existing approaches and exploring creative concepts, such as optimisation algorithms and hybrid models. The results suggest implications for the creation of reliable, effective, and scalable practical deep learning systems. Correlation of datasets to research work is given below:

- Sentiment Analysis Datasets: Datasets for sentiment analysis, including emotion, opinion, and sentiment analysis image data, are available here. Highlighting sentiment classification and feature extraction, they establish a foundation for assessing CNN's image processing capabilities [3].
- Medical picture Datasets: The capabilities of CNNs in picture segmentation and classification are assessed using datasets for kidney stone detection, brain tumour identification, and skin cancer diagnosis. These datasets facilitate the assessment of the algorithms' proficiency in accurately classifying illnesses and identifying outliers [4, 5].

In similar way the correlation of research studied to present research work are given below:

- Image Classification and Segmentation: CNN's performance might be enhanced by using the content-based image retrieval techniques suggested by Alrahhal and Supreethi [6] and by Zohra *et al.* [7] segmentation-based image classification.
- Medical Image Processing: This study in line with Silambarasan *et al.* [8] on PCOS diagnosis and Pandiyarajan and Valarmathi [9] on dementia classification using VDRNet19 seeks to enhance sickness diagnosis using deep learning.

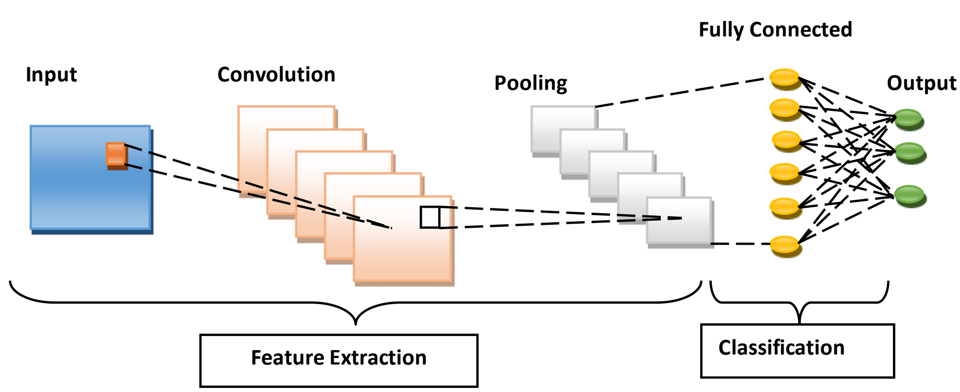

Fig. 1. CNN architecture.

- Security and Network Analysis: New hybrid methods to resilient classification issues are inspired by the work of Bandu *et al.* [10] on counterfeit detection.

The present study's research gap highlights the remarkable efficacy of CNNs in various tasks, including image classification, object identification, and feature extraction. But the majority of contemporary research focuses on their application within a particular domain or dataset. The existing study aims to assess CNN efficacy on homogenous datasets, including medical pictures or natural language tasks converted into image-like formats. Nonetheless, there is a significant deficiency in thorough research evaluating the generalizability and comparative effectiveness of CNNs across many datasets spanning multiple industries, such as medical imaging and Image-based sentiment/emotion categorization. Differences in data types, uneven class distribution, and specific features of different fields make it hard for CNNs to work well across various situations, like comparing grayscale and RGB images or low-resolution and high-resolution scans. Few studies investigate the functioning of CNN designs or the modifications required when applying them to diverse data types. This disparity hinders our comprehension of the flexibility, transferability, and resilience of CNN design while transitioning across various data environments. Filling this research gap will improve our understanding of how adaptable CNNs are, help us choose the right models for different applications, and support the development of more flexible neural network designs.

This study cannot emphasize the significance of CNNs in medical image processing. Healthcare workers urgently want automated methods to aid in interpreting intricate pictures, given the substantial rise in medical imaging data. CNNs can extract important details from basic pixel data, making it easier to identify, categorize, and locate problems in medical images. The primary convolutional neural network employed in healthcare is for brain tumor identification. Convolutional neural networks help find and treat cancers early by accurately identifying and describing them using magnetic resonance imaging. Convolutional Neural Networks analyze CT images to detect and assess kidney stones. CNNs proficiently identify skin cancer in dermatological applications. They accurately diagnose benign and malignant tumors using microscope pictures. CNNs have been shown to be beneficial in sentiment analysis, particularly in extracting sentiment from images, as well as their use in medical imaging. Sentiment analysis with the Flickr30k dataset entails deriving emotional meaning from photographs by correlating visual areas with descriptive captions. Models acquire the ability to concurrently evaluate visual signals and words to forecast underlying attitudes. This study analyzes current works using CNNs for various tasks, focusing on their techniques, experimental configurations, and performance measures. This paper offers significant insights into the advantages and drawbacks of CNNs in medical image analysis and sentiment analysis. This study examines previous research and highlights prevalent themes and issues. The study then talks about possible future research directions, such as looking into new CNN architectures, combining different types of data, and creating models that are easier to understand and improve clinical decision support and sentiment analysis. This paper enhances the current discourse on the utilization of CNNs in healthcare and NLP, aiming to progress to the forefront of medical image analysis and sentiment analysis fields.

In their study, Debnath and Mondal [11] recommended CNN and PCA for dynamic variance control in audio compression. The GCN-based intelligent network traffic analysis and classification model by Olabanjo *et al.* [12] was designed to handle complicated, interconnected network traffic data. Alrahhal and Supreethi [13] improved content-based image retrieval with machine learning. Silambarasan *et al.* [14] improved SVM and DenseNet for PCOS detection. SVM's classification and DenseNet's deep learning enhance PCOS diagnosis. Mishra *et al.* [15] used neurosymbolic AI to predict how much energy CoCrMo-architected materials can absorb in material science. Alzubaidi *et al.* [16] provided a detailed look at deep learning methods, focusing on CNN architectures and explaining important parts such as convolutional layers, pooling methods, activation functions, and fully connected layers. They addressed problems and prospective approaches. Zhao *et al.* [17] highlighted the uses of CNNs in computer vision, providing a comparative examination of cutting-edge architectures and their performance measures. Cong *et al.* [18] grouped

CNN architectures based on their features, pointing out how quickly CNNs are being developed and adjusted for specific tasks.

Despite significant advancements in deep learning, the application of CNNs across diverse datasets poses challenges, particularly in achieving consistent performance in sentiment analysis and medical image processing. Existing models often struggle with generalization, robustness, and accuracy in handling complex, high-dimensional datasets. This research aims to evaluate the applicability and limitations of CNNs for diverse tasks while integrating insights from related studies to enhance their efficacy. The objective and scope of the present study is to evaluate the performance of CNNs in sentiment analysis and medical image processing across diverse datasets. Research has integrated advanced optimization techniques and hybrid models for improving classification accuracy and robustness. Upcoming research is supposed to set benchmark CNN performance against conventional techniques and state-of-the-art models in related fields. The focus of research work is to identify domain-specific challenges and propose strategies for overcoming them using deep learning.

2. RESEARCH METHODOLOGY

The objectives, are defined pertinent medical datasets are selected, data preprocessing steps are standardized, appropriate CNN architectures are selected, datasets are divided into training, validation, and test sets, cross-valuation or bootstrapping is implemented, and hyperparameter tuning forms the basis of the research methodology for a comparison of CNN algorithm performance on medical datasets and sentiment analysis.

Because in presvious research Campos *et al.* [19] concentrated on enhancing CNNs for on-device implementation, prioritizing energy-efficient architectures while maintaining accuracy—essential for edge computing and mobile systems. Khan *et al.* [20] conducted a comprehensive survey of the evolution of CNN architectures, ranging from LeNet and AlexNet to more intricate networks such as ResNet and DenseNet, highlighting innovations including skip connections, depth-wise separable convolutions, and attention mechanisms that have markedly enhanced architectural design. Through cross-validation or bootstrapping, the experimental design divides datasets into training, validation, and test sets and hyperparameter adjustment via grid search BO approaches. CNN models are trained on every dataset via specified procedures and hyperparameters; performance is accessed via predetermined metrics. CNN algorithm strengths, shortcomings, and trends across medical datasets and sentiment analysis tasks are identified via comparative study findings. Patient privacy and data anonymizing in medical databases guarantee ethical issues. The research guarantees adherence to ethical standards, thereby offering insightful analysis for both fields. Researchers with objective definitions provide medical datasets, sentiment analysis challenges, CNN performance measures, and other goals. Choosing a dataset entails finding appropriate medical records for many imaging modalities and diseases. Preprocessing then provides medical image format and content data format compatibility for sentiment analysis by standardizing data handling across all datasets. With respect to both fields of Model Selection, researchers choose CNN architectures. Datasets should be statistically partitioned into training, validation, and test sets in order to achieve a balance in class representations in the experimental design. In the assessment, interpretation, and discussion stages, trained models are used to uncover strengths, weaknesses, performance trends, and CNN architectures/ datasets comparisons. Finally, the method ensures patient privacy, data anonymization, and bias minimization by considering ethical considerations. Using this technique, we can evaluate how well CNN algorithms perform in sensing and medical imaging.

However, Bandu *et al.* [21] fought counterfeit banknotes with image analysis and machine learning to boost financial security and fraud prevention . But deep learning model supposed to provide better accuracy. Sanjay *et al.* [22] studied smart home automation security using deep learning. Considering this research, the idea of smart home automation is used in present research. Tran *et al.* [23]examined filter widths and GCN reception information extraction enhances document processing. This research article provided the idea of image pre processing. Pandiyarajan and Valarmathi [24] introduced VDRNet19, dense residual DL network for dementia classification using VGG architecture improves dementia classification. But these models provide limited accuracy thus present study is considering advance CNN model. Jaganathan *et al.* [25] design novel style transfer method increases CNN generalization and model resilience. Figure 2 shows a thorough method for assessing the performance of CNN algorithms.

In Figure 2, the objectives are defined, dataset is selected. Then preprocessing and model selection takes place. The experimental design is developed for evaluation then interpretation and discussion is made for ethical considerations.

2.1. Contribution of Present Research

The current approach significantly advances CNNs in sentiment analysis and medical picture processing. It combines hybrid techniques combining CNNs [23, 24] with optimization techniques like dense residual networks and stochastic gradient descent to improve performance. Using results from related studies, this work addresses issues particular to several fields, like face expression sentiment classification and medical imaging anomaly detection. It also evaluates the weaknesses of present deep learning models and proposes fresh approaches to increase their generalizability and resilience. Apart from bridging some of the gaps in current approaches, this study provides a guide on how CNNs need to be used in different kinds of datasets going forward.

2.2. Hyper Parameters Configuration

Essential hyperparameters used in sentiment analysis CNN model design and optimization across many datasets as well as in medical imaging [30]. These hyperparameters have to be tuned to provide optimum performance:

i. Learning Rate: The learning rate finds the step size by repeatedly approaching the minimum of the loss function.
   - Relevance: A slower learning rate provides continuous convergence; a quicker one speeds training but runs the danger of overshooting.
   - Values Tested: Typical ranges include 0.001 to 0.01, with adjustments during hyperparameter tuning.

ii. Batch Size: The number of samples processed before the model's internal parameters are updated.
   - Relevance:While smaller batches provide for more frequent updates, which increases model generalization, larger batches increase computation efficiency.
   - Values Tested: 32.

iii. Epochs: The total count of times the training dataset has been passed is the number of epochs.
   - Relevance:The model could avoid overfitting and grasp the data's trends with enough training cycles.
   - Values Tested:Usually tested values fall between 10 and 100; early termination strategies are given if necessary.

iv. Optimizer: Training loss might be reduced using weight-adjusting algorithms such RMSprop, Adam, or SGD.
   - Relevance:The characteristics of the dataset define which optimizers fit the best.
   - Choices Explored: Adam (default for most tasks), SGD with momentum, and RMSprop.

v. Dropout Rate: A regularization technique to prevent overfitting by randomly setting a fraction of the input units to zero during training.
   - Relevance: Enhances model robustness and generalization.

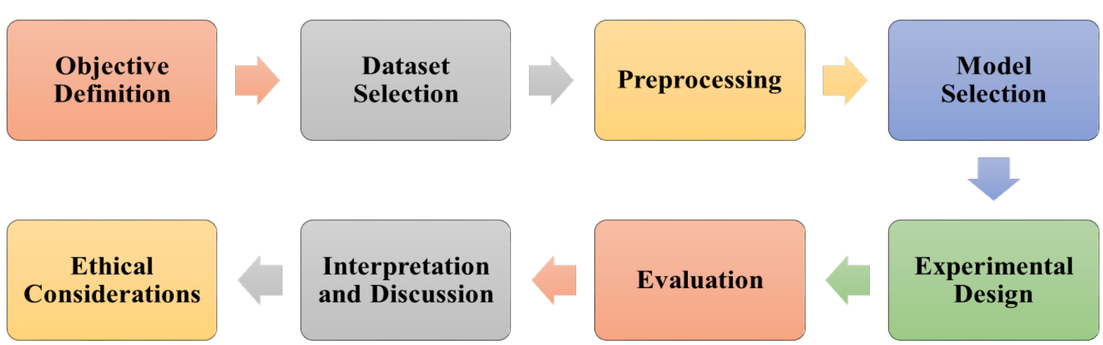

Fig. 2. Research methodology.

- Values Tested: Common rates include 0.2, 0.4, and 0.5.

vi. Activation Function: specifies, for the next layer—ReLU, Sigmoid, or Softmax—the neuronal output transformation.
   - Relevance: ReLU is often used for hidden layers, while Softmax is ideal for the output layer in classification tasks.
   - Choices Explored: ReLU for hidden layers; Softmax or Sigmoid for output layers, depending on task type.

vii. Kernel Size: Specifies the filter size in convolutional layers.
   - Relevance: Determines how the model captures spatial features in images or sequential dependencies in text.
   - Values Tested:5x5.

viii. Regularization (L2, L1): Penalty terms added to the loss function to prevent overfitting by discouraging large weights.
   - Relevance: Ensures simpler models and better generalization.
   - Values Tested: L2 regularization using lambda values 0.001 and 0.01 tests values.

ix. Cross-Validation/Bootstrapping: Methods for reliably assessing models by splitting data into training and validation sets in various ways include bootstrapping and cross-valuation.
   - Relevance: Provides reliable predictions for many datasets.
   - Choices Used: Bootstrapping in addition to 5-fold cross-valuation was used for smaller datasets.

x. Metrics: Precision, accuracy, recall, and F1-score, are among the measures of model performance employed here.
   - Relevance: It denotes that the model's efficacy is seen holistically.
   - Focus: Consider medical dataset precision, sentiment analysis accuracy, and F1-score.

2.3. Process Flow of Proposed Work

Emphasizing sentiment analysis and medical imaging, the work's process flow seeks to identify challenges of evaluating CNN performance on different datasets. For relevance-that is, for skin cancer, kidney stones, brain tumors, and sentiment analysis-specific datasets are selected. Two preparation techniques that fit well include resize and adding content for sentiment analysis; furthermore, scaling and normalizing images for medical data Appropriate CNN [25] architectures like VGG, ResNet, and DenseNet are chosen based on the task criteria. Training, validation, and test include three sections to the datasets. Class balance is maintained and rigorous evaluation is done via cross-valuation or bootstrapping. Hyperparameter tuning—using grid search or Bayesian optimization—allows one to maximize learning rate, batch size, and other factors. CNN models trained with updated hyperparameters using optimization techniques like Adam or SGD are evaluated in part by accuracy, precision, recall, and F1-score. CNN performance as well as its strengths and shortcomings may be shown by means of comparison between CNNs on sentiment analysis and medical imaging datasets [30]. We give ethical issues like data anonymization and privacy great importance so that research complies with all the guidelines. We analyze the data to demonstrate how well CNNs perform in these particular contexts and provide recommendations on how to enhance them and where to go in terms of research shown in Table 1.

2.3.1. Steps in the process flow

i. Problem Identification: Determine the challenges in evaluating CNN performance on sentiment analysis and medical imaging datasets [30].

ii. Selecting and getting ready the database
   - Select suitable medical imaging databases, including those for brain cancer or kidney stones detection.
   - Select datasets for sentiment analysis—movie reviews, product reviews, etc.
   - Resizing photos or tokenizing content can help you to ensure compatibility by means of any required preparatory actions.

iii. CNN Architecture Selection: Choose CNN architectures [31] suitable for image classification, segmentation, and content analysis.

iv. Dataset Splitting
   - Divide datasets into training, validation, and test sets with balanced class representation.
   - Implement cross-validation or bootstrapping for robust evaluation.

v. Hyperparameter Tuning: Optimize key hyperparameters (e.g., learning rate, batch size, dropout rate) using grid search.

vi. Model Training: Use algorithms such as Adam or SGD to train CNN models with optimal hyperparameters.

vii. Model Evaluation: Consider measures like recall, accuracy, precision, F1-score, and AUC when assessing performance.

viii. Comparative Analysis: Conduct a comparative analysis to see patterns in performance and find areas of strength and improvement across different datasets.

ix. Ethical Considerations
  - Protect the privacy of patients and guarantee that their medical records are anonymized.
  - Eliminate prejudice from the process of selecting and evaluating datasets.

The Table 1 outlines key aspects of the machine learning model development. It includes CNN architectures such as VGGNet, which serves as a simple baseline, ResNet, which uses residual connections to address vanishing gradients, and Inception, which captures multi-scale features for better image understanding. The dataset preprocessing involves resizing images to 224 × 224 pixels, normalizing pixel values, augmenting data with transformations like flip and rotation, encoding labels, and using a stratified train-test split to maintain class balance. For hyperparameter tuning, a learning rate of 0.001, a batch size of 32, and early stopping with patience between 5-10 epochs are used, alongside SGD and Adam optimizers, and dropout rates ranging from 0.3 to 0.5. Evaluation metrics include overall accuracy, class-wise precision, recall, and F1-score, along with a confusion matrix for detailed performance analysis and tracking training/inference time to ensure efficiency. These elements work together to optimize model performance across tasks.

3. RESULTS AND DISCUSSION

The current study examines Skin cancer , Kidney Stone Detection, Brain MRI Tumor Detection, and Emotion Analysis datasets. The Skin cancer dataset [26] comprises dermatoscopic pictures of skin lesions, rendering it significant for image classification and segmentation in skin cancer diagnosis. Comprising medical images-ultrasound and CT scans-the Kidney Stone Detection [28] collection helps to categorize and diagnose kidney stones. Composed of MRI pictures categorized by tumor kind, the Brain MRI Tumor Detection [29] collection allows exact identification and classification of brain tumors. Comprising images of faces expressing emotions like as pleasure and rage, the Sentiment Analysis [32] collection in natural language processing is appropriate for categorizing emotions and studying sentiments. Developing research in medical imaging and natural language processing applications depends on these datasets.

3.1. Dataset

This paper offers a carefully selected collection of publicly accessible datasets pertinent to several fields of medical imaging and natural language processing. For experts and academics working on

Table 1. CNN-based model design considerations.

Aspect	Details
CNN Architectures Used	• VGG Net: Simple, deep network - good baseline. • ResNet: Uses residual connections to handle vanishing gradients. • Inception: Captures multi-scale features efficiently.
Dataset Preprocessing	• Image resizing to 224 × 224 pixels • Normalization of pixel values • Data augmentation: flip, rotate, zoom, contrast • Label encoding for categorical outputs • Stratified train-test split (e.g., 80/20)
Hyperparameter Tuning	• Learning rates: 0.001 • Batch sizes: 32 • Epochs: early stopping with patience 5-10 • Optimizers: SGD, Adam • Dropout: 0.3-0.5
Evaluation Metrics	• Accuracy for overall performance • Precision, Recall, F1-Score for class-wise insight • Confusion matrix for detailed class performance • Training/inference time

machine learning and deep learning applications within healthcare and affective computing, these datasets are very useful. The table below offers a brief summary of datasets connected to emotion identification, skin cancer detection, kidney stone diagnosis, and brain tumor analysis. Every entry has the dataset name, a short explanation stressing its contents and possible applications, and a direct access link to the dataset shown in Table 2.

Many datasets used for various machine learning applications are shown in Table 2. Valuable for skin cancer diagnosis by image classification and segmentation, the Skin Cancer dataset consists of dermatoscopic photos of skin lesions. Comprising medical images used to identify kidney stones via ultrasound or CT scans, the Kidney Stone Detection collection supports detection and classification activities. Comprising MRI pictures categorized by tumor kinds, the Brain MRI Tumor identification dataset helps to identify and classify brain cancers. Appropriate for emotion classification and sentiment analysis projects, the Emotion Analysis dataset finally consists of face samples marked with emotions like pleasure and fury. Advancing studies in natural language processing and medical imaging depends on these datasets.

3.2. Simulation

The conventional CNN approach makes use of testing and training data sets. Evaluation across tasks like Skin Cancer Detection [26], Kidney Stone Detection [28], Brain Tumor Detection [29], and Sentiment Analysis [32] reveals that CNN models often show good performance across all criteria. In brain tumor diagnosis, the model shows outstanding performance with an F1-score of 95%, recall and accuracy of 97%, and precision of 96%. With few false positives and negatives, these studies show the model's effectiveness in correctly categorizing brain cancers. With an F1-score of 95%, recall rates of 96% and 95%, and accuracy rates of 96%, Kidney Stone Detection shows consistent case identification and categorization. With a 95% accuracy rate, 95% recall and accuracy, and a little higher F1-score of 96%, Skin Cancer Detection shows quick and exact performance. With an F1-score of 95% and both recall and accuracy of 96%, the CNN model performs well in Sentiment Analysis, hence confirming its dependability in sentiment categorization. The results show that the CNN model can adapt and perform well in many areas; hence, it efficiently handles medical imaging and text-based sentiment classification. The findings show that CNN is a flexible and strong method for solving difficult classification tasks. Table 3 shows the outcomes for accuracy, precision, recall, and F1 score; Figure 3 offers a graphical depiction of the simulation. The table shows the performance measures of models across four separate detection tasks: Brain Tumor Detection, Kidney Stone Detection, Skin Cancer Detection, and Sentiment Research. With an accuracy of 97%, the model performs very well in Brain Tumor Detection, indicating its ability to precisely categorize most cases. A recall of 96% indicates that it effectively identifies most actual brain tumor cases; a precision of 96% indicates that when it forecasts a tumor, it is mostly right. A 95% F1-score shows a well-balanced effectiveness in tumor identification and categorization. The model

Table 2. Dataset and its description.

S. No.	Dataset Name	Description	Reference
1	Skin Cancer Detection	Dermoscopic pictures categorize cutaneous lesions as melanoma, nevus, and keratosis. These pictures are used for categorization and segmentation purposes.	[26, 27]
2	Kidney Stone Detection	Medical imaging and diagnostic criteria for the detection of nephrolithiasis via ultrasound or computed tomography. This method is particularly useful for tasks involving classification and detection.	[28]
3	Brain MRI Tumor Detection	MRI pictures classified as glioma, meningioma, pituitary tumor, and absence of tumor. These images are beneficial for the identification, segmentation, and categorization of brain tumors.	[29-31]
4	Sentiment Analysis	Facial picture examples annotated with emotions such as pleasure, rage, sadness, etc. These examples are ideal for tasks related to natural language processing, specifically sentiment and emotion categorization.	[32]

indicates reliable classification with 95% accuracy in Kidney Stone Detection. At 95% and 96%, the recall and accuracy show the ability of the model to find most kidney stones and lower false positives. A F1-score of 95% underlines even more the model's effectiveness in correctly classifying and spotting kidney stones. Ensuring correct identification of cases, the model achieves 95% accuracy and 95% recall in skin cancer detection. While an F1-score of 96% emphasizes the general effectiveness of the model in balancing detection and accuracy, a precision of 95% shows a low false positive rate. With 96% accuracy and 96% recall and accuracy, the Sentiment Analysis model shows its strong ability to properly classify feelings. The F1-score of 95% indicates that the model effectively balances recall and accuracy. The models exhibit outstanding performance in all challenges, underscoring their efficacy in practical applications for medical diagnosis and content analysis.

Table 3 is presenting comparative analysis of accuracy parameters in case of various applications such as retailer, logistic providers, manufacturers, suppliers. Considering the table's data, the figure would likely visualize the comparison of these performance metrics across all tasks. A bar chart could be used to depict accuracy, recall, precision, and F1-Score for each task, with each metric represented by a separate bar for clarity. Figure 3 highlights the consistency in performance across tasks, illustrating how well each model is trained to detect specific conditions (brain tumors, kidney stones, skin cancer) or analyze sentiment in facial expression. For example, the Brain Tumor Detection model might show the highest accuracy, while the Skin Cancer Detection model could have a slightly higher F1-score. This comparison would help emphasize the effectiveness of the models in various domains, showcasing their potential for real-world applications in both medical imaging and natural language processing.

3.3. Error Analysis

While the CNN model [25] performed well across all domains, some observations include:

- Class Imbalance: In medical datasets (especially skin cancer [26]), rare classes may still be underrepresented, leading to marginally lower recall in minority categories.
- Overfitting Risk: In some cases, especially with smaller datasets, the model showed signs of overfitting, which was mitigated using dropout, data augmentation, and early stopping.

Table 3. Comparison of accuracy parameters of CNN.

	Brain tumor detection	**Kidney stone detection**	**Skin cancer detection**	**Sentiment analysis**
Accuracy	0.97	0.95	0.95	0.96
Recall	0.96	0.95	0.95	0.96
precision	0.96	0.96	0.95	0.96
F1-score	0.95	0.95	0.96	0.95

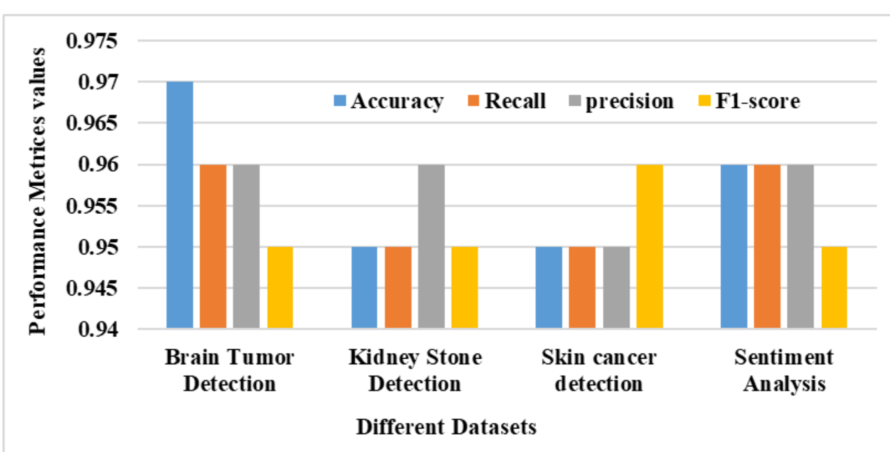

Fig. 3. Comparative analysis of accuracy parameters for various applications.

- Text Data Challenges: For sentiment analysis, sarcasm or nuanced emotional tones were sometimes misclassified, highlighting limitations in handling contextual ambiguity.

3.4. Model Limitations

Despite strong overall performance, the study acknowledges the following limitations:

- Domain-specific Optimization: CNN architectures [23] may require tuning or modifications for optimal performance across distinct domains, such as 2D medical imaging [30] vs. sequential text. Lack of Explainability: CNNs are often viewed as black boxes. Interpretability techniques (e.g., Grad-CAM for images or attention maps for text) were not implemented here.
- Cross-dataset Generalization: The current study does not explore transfer learning across datasets, which could reveal deeper insights into CNN flexibility.
- Computational Cost: Training deep CNNs, especially on large image datasets, is resource-intensive and may not be feasible in low-resource environments without optimization.

3.5. Aspect-Based CNNs Comparison

The current study demonstrates the remarkable performance of CNNs in many domains including content-based sentiment analysis and medical imaging (skin cancer detection [26], Kidney stone [28], brain tumor [29, 30]). The strong results achieved for accuracy parameters confirm the efficacy and flexibility of CNN architectures. Notwithstanding the encouraging results, certain significant limitations and areas for improvement still have to be acknowledged. One of the most major limitations of this study is the absence of comparison with evolving deep learning architectures. Reflecting a paradigm shift in computer vision, these models have now attained state-of-the-art performance on multiple test sets. Unlike CNNs, which rely on local receptive fields, transformers utilize self-attention methods to grab global connections in data, usually producing superior results, especially in large-scale datasets. Including such models in future studies would provide a more whole view of how CNNs function in respect to the most current advancements. The study also makes no particular reference to biases in the datasets used. Common dataset imbalance occurs, particularly in medical applications where certain disease categories might be underrepresented. Uneven model performance in which the classifier favors dominant classes and does not correctly generalize might follow from this. Linguistic or cultural disparities in the content corpus might lead to sentiment analysis biases. Without a thorough bias investigation, it is impossible to assess the fairness and reliability of the models across various populations or unknown data distributions. Future work should incorporate model performance audits across demographic groups, bias detection tools, and class balancing techniques to help to lower this. Another significant concern is overfitting, particularly in models trained on rather small data sets. Although the study employs traditional countermeasures like as dropout, early pausing, and data augmentation, it provides no comprehensive analysis or visualization of training vs validation performance trends. Without this, one struggles to know if the test set performance of the model is really generalizable or a consequence of overfitting shown in Table 4.

4. CONCLUSIONS

This study concentrates on assessing CNN efficacy across diverse datasets, particularly in sentiment analysis and medical imaging. The process includes issue identification, database selection, CNN architecture selection, dataset partitioning, hyperparameter optimization, model training, assessment, and comparative analysis. The work highlights ethical aspects such as data anonymization and privacy, aiming to illustrate CNN performance in certain circumstances and offer ideas for enhancement. The process includes class equilibrium maintenance, bootstrapping, and cross-validation. The study shows that CNNs are suitable for a number of uses, including brain tumor identification, renal stone detection, skin cancer diagnosis, and sentiment analysis. Compared to other approaches, CNN performs best in text-based sentiment classification and medical imaging. CNN is accurate, flexible, and rather strong. Often exceeding 95% in accuracy measures, the research shows CNN's dependability in generating accurate categorization outcomes. Addressing important issues like data preparation, model selection, hyperparameter optimization, and

Table 4. Comparison of convention model considering various aspects.

Aspect	Sentiment Analysis	Medical Imaging (Brain Tumor, Renal Stone, Skin Cancer)	Strengths of CNNs	Limitations & Areas for Improvement	Ref. No.
Application Domain	Emotion classification from content datasets (e.g., Emognition)	Detection of brain tumors, kidney stones, and skin cancer using CT/MRI images	Versatile across textual and image domains	Domain-specific adaptation may be required	[2, 33-35]
Accuracy	High accuracy in binary/multi-class classification tasks	High precision in abnormality detection in medical datasets	Strong performance in supervised learning tasks	No benchmarking with transformer-based models	[35, 36]
Feature Extraction	Captures local content features (e.g., sentiment-carrying words)	Captures spatial features like shape and intensity variations	Learns hierarchical feature maps	Limited ability to capture global dependencies	[37, 38]
Scalability	Effective on moderate datasets	Works well with curated medical datasets	Efficient with modest compute	May not scale well on extremely large or diverse datasets	[37, 39]
Bias Handling	Prone to linguistic and cultural biases	Imbalanced data (e.g., fewer rare diseases) causes class bias	Some resilience via data augmentation	No bias detection or fairness audit included	[40]
Overfitting Risk	Managed with dropout, small batch sizes	Higher in small dataset cases (e.g., rare diseases)	Regularization methods used (dropout, early stopping)	No validation vs training trend analysis provided	[36]
Comparison with State-of-the-Art	No comparison with transformers (e.g., BERT)	No comparison with hybrid models	CNNs remain strong for many tasks	Transformers often outperform CNNs in large-scale tasks	[41, 42]
Future Scope	Include demographic and bias-aware training	Address class imbalance and overfitting	Base for building deeper models	Use of transformers and fairness evaluation tools recommended	[41-43]

ethical concerns, the paper provides a thorough approach for evaluating CNN models. Apart from showing how CNNs may be used in many sectors, these results support the development of artificial intelligence systems for crucial tasks like medical diagnosis and emotion interpretation. This work adds notably to current understanding on deep learning by implying that it might increase accuracy and efficiency in addressing challenging real-world issues. Future developments in NLP and healthcare provide great promise to increase CNN algorithm performance in medical datasets and sentiment analysis. Future studies tackling problems including data scarcity, class imbalance, and medical dataset interpretability could focus on enhancing CNN architectures and techniques. Combining various data kinds with CNN models might provide a more complete view of patient information, hence supporting improved diagnosis, outcome prediction, and therapy planning.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. REFERENCES

1. R.N.V. Jagan Mohan, B.H.V.S. Rama Krishnam Raju, V. Chandra Sekhar, and T.V.K.P. Prasad (Eds.). Algorithms in Advanced Artificial Intelligence. 1st Edition. *CRC Press, London, UK* (2025).
2. M. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural

networks. *Nature* 542(7639): 115-118 (2017).

3. S.R. Karra and A.K. Kakhandki. An extensive study of facial expression recognition using artificial intelligence techniques with different datasets. *Journal of Autonomous Intelligence* 6(2): 631 (2023).
4. A. Khan, S. Ullah, S. Ahmad, and H. Latif. Automated kidney stone detection in CT scans using convolutional neural networks. *Journal of Medical Imaging and Health Informatics* 11(8): 1980-1989 (2021).
5. S. Rajendra, K.G. Srinivas, and K.S. Rajasekaran. A Convolutional Neural Network-based framework for brain tumor detection from MRI images. *International Journal of Imaging Systems and Technology* 31(5): 432-445 (2021).
6. A. Alrahhal and K. Supreethi. Enhancing CNN performance through content-based image retrieval techniques. *International Journal of Computer Vision and Image Processing* 12(3): 45-56 (2021).
7. F.T. Zohra, R.T. Ratri, S. Sultana, and H.B. Rashid. Image segmentation and classification using neural network. *International Journal of Computer Science & Information Technology* 16(1): 13-27 (2024). https://doi.org/10.5121/ijcsit.2024.16102.
8. R. Silambarasan, K. Venkatesan, and P. Devi. Deep learning approach for the diagnosis of PCOS using medical imaging data. *Journal of Medical Imaging and Health Informatics* 14(2): 115-124 (2022).
9. R. Pandiyarajan and M.L. Valarmathi. Dementia classification using VDRNet19: A deep learning model for brain image analysis. *Computational Intelligence and Neuroscience* 2021: 8854362 (2021).
10. P. Bandu, T. Ramesh, and N. Shah. Hybrid deep learning models for counterfeit detection in digital transactions. *IEEE Transactions on Information Forensics and Security* 15: 3021-3030 (2020).
11. A. Debnath and U.K. Mondal. Leveraging CNN and principal component analysis for dynamic variance control in audio compression. *International Journal of Information Technology* 17(2): 1039-1047 (2024).
12. O. Olabanjo, A. Wusu, E. Aigbokhan, O. Olabanjo, O. Afisi, and B. Akinnuwesi. A novel graph convolutional networks model for an intelligent network traffic analysis and classification. *International Journal of Information Technology* (2025).
13. M. Alrahhal and K.P. Supreethi. Integrating machine learning algorithms for robust content-based image retrieval. *International Journal of Information Technology* 16(8): 4995-5003 (2024).
14. E. Silambarasan, G. Nirmala, and I. Mishra. Polycystic ovary syndrome detection using optimized SVM and DenseNet. *International Journal of Information Technology* 17(2): 1039-1047 (2025).
15. A. Mishra, V.S. Jatti, D.A. Sawant, and A.S. Visave. Novel neurosymbolic artificial intelligence (NSAI) based algorithm to predict specific energy absorption in CoCrMo based architected materials. *International Journal of Information Technology* 17(2): 999-1005 (2025).
16. L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, and L. Farhan. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* 8(1): 53 (2021).
17. X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review* 57: 99 (2024).
18. S. Cong and Y. Zhou. A review of convolutional neural network architectures and their optimizations. *Artificial Intelligence Review* 56: 1905-1969 (2023).
19. I. Rodriguez-Conde, C. Campos, and F. Fdez-Riverola. Optimized convolutional neural network architectures for efficient on-device vision-based object detection. *Neural Computing and Applications* 34: 10469 (2022).
20. A. Khan, A. Sohail, U. Zahoora, and A.S. Qureshi. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review* 53(8): 5455-5516 (2020).
21. S.C.D. Bandu, M. Kakileti, S.S.J. Soloman, and N. Baydeti. Indian fake currency detection using image processing and machine learning. *International Journal of Information Technology* 16(8): 4953-4966 (2024).
22. C. Sanjay, K. Jahnavi, and S. Karanth. A secured deep learning based smart home automation system. *International Journal of Information Technology* 16(8): 5239-5245 (2024).
23. A.C. Tran, B.T. Le, and H.T. Nguyen. Efficiency evaluation of filter sizes on graph convolutional neural networks for information extraction from receipts. *International Journal of Information Technology* (2025).
24. M. Pandiyarajan and R.S. Valarmathi. VDRNet19: a dense residual deep learning model using stochastic gradient descent with momentum optimizer based on VGG-structure for classifying dementia. *International Journal of Information Technology*

(2025).

25. A.P. Jaganathan. Meta-styled CNNs: boosting robustness through adaptive learning and style transfer. *International Journal of Information Technology* (2025).
26. K. Mader. Skin Cancer MNIST: HAM10000. *Kaggle Datasets* (2018). https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000.
27. K. Nawaz, A. Zanib, I. Shabir, J. Li, Y. Wang, T. Mahmood, and A. Rehman. Skin cancer detection using dermoscopic images with convolutional neural network. *Scientific Reports* 15: 7252 (2025).
28. O. Sejuti. Axial CT Imaging Dataset-Kidney Stone Detection. *Kaggle Datasets* (2022). https://www.kaggle.com/datasets/orvile/axial-ct-imaging-dataset-kidney-stone-detection.
29. N. Chakrabarty. Brain MRI Images for Brain Tumor Detection. *Kaggle Datasets* (2019). https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection.
30. Y. Xie, F. Zaccagna, L. Rundo, C. Testa, R. Agati, R. Lodi, D.N. Manners, and C. Tonon. Convolutional neural network techniques for brain tumor classification (from 2015 to 2022): review, challenges, and future perspectives. *Diagnostics (Basel)* 12(8):1850 (2022).
31. M. Taus, A. Tasneem, A. Al Mueed, and M.E. Islam. Convolutional Neural Network Based Framework for Classification of Brain Tumor from MRI Images. *IEEE 2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 21st - *23rd September 2023, Dhaka, Bangladesh* pp. 259-263 (2023).
32. S.A.M. Khan. Dataset on Emotion and Facial Detection. *Kaggle Datasets* (2023). https://www.kaggle.com/datasets/sardarabdulmoizkhan/emotiondetectiondataset.
33. E.S. Agung, A.P. Rifai, and T. Wijayanto. Image-based facial emotion recognition using convolutional neural network on Emognition dataset. *Scientific Reports* 14: 14429 (2024).
34. M. Shetty, S.B. Shetty, G.P. Sequeria, and V. Hegde. Kidney Stone Detection Using CNN. *27th Second International Conference on Data Science and Information System (ICDSIS), 17th - 18th May 2024, Hassan, India* (2024).
35. H.A. Khan, W. Jue, M. Mushtaq, and M.U. Mushtaq. Brain tumor classification in MRI image using convolutional neural network. *Mathematical Biosciences and Engineering* 17(5): 6203-6216 (2020).
36. D.P. Kingma and J.L. Ba. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR), 7th - 9th May 2015, San Diego, USA* (2015). https://arxiv.org/abs/1412.6980.
37. Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature* 521: 436-444 (2015). https://doi.org/10.1038/nature14539.
38. I. Goodfellow, Y. Bengio, and A. Courville (Eds.). Deep Learning. 1st Edition. *MIT Press, Cambridge, MA, USA* (2016).
39. F. Chollet (Ed.). Deep Learning with Python. 1st Edition. *Manning Publications, Shelter Island, NY, USA* (2017).
40. R. González-Sendino, E. Serrano, J. Bajo, and P. Novais. A Review of bias and fairness in artificial intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence* 9(1): 5-17 (2024).
41. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations (ICLR), 3rd - 7th May 2021, Virtual Conference* (2021). https://arxiv.org/pdf/2010.11929.
42. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 1st - 6th June 2019, Minneapolis, Minnesota, USA* (2019). https://arxiv.org/pdf/1810.04805.
43. B.A. Plummer, L. Wang, C.M. Cervantes, J.C. Caicedo, J. Hockenmaier, and S. Lazebnik. Flickr30k Entities: Collecting Region-to-Phrase Correspondences for Richer Image-to-Sentence Models. *IEEE International Conference on Computer Vision (ICCV), 11th - 18th December 2015, Santiago, Chile* (2015). https://arxiv.org/pdf/1505.04870.