OPEN ACCESS JOURNALS
Indiana Journal of Multidisciplinary Research
Abbreviated Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.21610415

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Articles

Volume-06|Issue-04|2026

Systematic Framework for Pulmonary Disease Evaluation

Dr. B. P. Pradeep Kumar[1] and Manjunath K. M[2]

Professor, Computer Science and Design, Atria Institute of Technology, Bengaluru, Karnataka, India.
Assistant Professor, Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, Karnataka, India.

Article History
Received: 05.05.2026
Accepted: 22.06.2026
Published: 25.07.2026

Citation
Kumar, B. P. P. & Manjunath, K. M. (2026) Systematic Framework for Pulmonary Disease Evaluation. *Indiana Journal of Multidisciplinary Research, 6*(4), 414-420.

Abstract: The necessity of automatically identifying anomalies in CT scans is highlighted by the rising incidence of lung ailments brought on by poor air quality. Through the analysis of large datasets, machine learning is essential for the effective detection of diseases like lung cancer through lung nodule assessment. Usually, there are three steps in the diagnostic process: pre-processing the images, using deep learning (DL) models, and deploying convolutional neural networks (CNNs). While DL algorithms extract essential information for precise classification, pre-processing improves the clarity of CT images. CNNs are especially good at seeing intricate patterns that enable accurate differentiation between tissues that are healthy and those that are not. The accuracy and efficiency of lung disease diagnostics are greatly increased by this integrated approach, which promotes quicker detection, better treatment planning, and better patient outcomes.

Keywords: Neural networks, machine learning (ML), CNN, CT scans, lung cancer.

INTRODUCTION

Lung cancer is among the most deadly and pervasive malignancies in the world. It is characterized by uncontrolled cell growth in lung tissues. This malignant process interferes with normal cellular functions, creating tumors that compromise the respiratory system and neighboring organs. The disease primarily occurs in two forms: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). Timely identification is crucial for successful treatment; however, the disease often progresses silently in its initial phases, resulting in delayed diagnosis. Symptoms such as chronic cough, pain in the chest, alterations in voice, and breathing difficulties typically emerge in advanced stages.

According to the American Cancer Society, in the United States, lung cancer was the second most common kind of cancer as of 2022, with projections of 236,740 new cases and 130,180 fatalities. Tobacco use is the primary risk factor, followed by exposure to radon gas, asbestos, diesel fumes, air pollutants, and certain toxic metals including arsenic and chromium [21]. Moreover, occupational risks in sectors such as construction, rubber manufacturing, and painting increase susceptibility [22]. Genetic factors and medical history, including tuberculosis, also contribute to increased vulnerability.

Given the complexity of lung cancer diagnosis and the challenges of early detection, advanced technology techniques are essential. This study suggests using machine learning (ML) methods—more especially, Convolutional Neural Networks (CNNs)—to segment and analyze computed tomography (CT) scan pictures in order to make the accuracy and efficiency of lung cancer detection **better** [23].

CT scans serve as a vital diagnostic tool, providing detailed cross-sectional images of lung structures that assist in identifying abnormalities. However, manual interpretation of these images can be time-intensive and error-prone. CNNs, with their capacity to recognize complex patterns in images through hierarchical feature extraction, offer a powerful method for automating and improving diagnostic accuracy [24].

The proposed methodology encompasses several stages: enhancing CT image clarity through preprocessing, training CNN models to identify cancerous patterns, and employing these models for classification and detection of abnormalities. This automated approach reduces the potential for human error, facilitates large-scale image analysis, and optimizes the diagnostic process [25].

This study highlights the potential of incorporating AI-driven techniques into medical diagnostics, giving new instruments for the early diagnosis of lung cancer [26]. Such developments have the potential to significantly enhance patient outcomes and survival rates by enabling prompt diagnosis, therefore tackling a significant healthcare burden.

***Corresponding Author:** Dr. B. P. Pradeep Kumar

MATERIALS AND METHODS

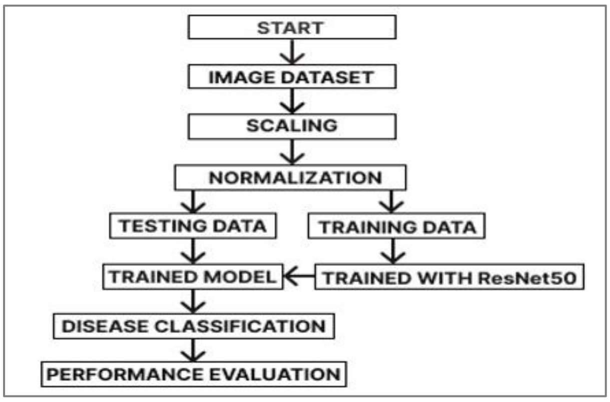

Fig. 1. Block Diagram

Figure 1 illustrates the development of an effective machine learning (ML) model for detecting lung cancer, which starts with assembling a comprehensive and trustworthy dataset. The accuracy and resilience of the model depend on proper data collection, organization, and preprocessing. This research utilizes CT scan images from public sources and employs systematic preprocessing methods to prepare the dataset for deep learning applications.

This study employs a dataset of CT scan images acquired from Kaggle, a reputable platform for well-curated, publicly available datasets. The collection consists of 1,097 images classified into three categories, as shown in Figure 2: benign, malignant, and normal (non-cancerous). These images are provided in common formats such as JPG and PNG, ensuring compatibility with various machine learning (ML) frameworks. Each image underwent careful review to verify accurate classification and appropriate representation. The dataset's quality and reliability form the basis for subsequent preprocessing and model development. These steps are vital in creating a model that can effectively identify patterns and differentiate between cancerous and non-cancerous conditions.

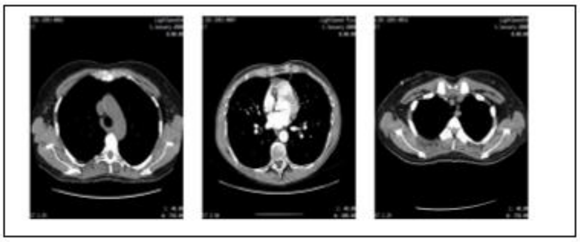

Fig. 2. CT scan input pictures with nodule types: (a) Benign, (b) Malignant, (c) Normal

After dataset collection, the next phase involved scaling the images to ensure uniformity and compatibility with deep learning models. This step optimizes the dataset for computational efficiency while maintaining essential features necessary for accurate diagnosis. Images were resized to a standard dimension of 64 × 64 pixels to achieve uniformity across the dataset. This resolution was selected because it strikes a balance between computational efficiency and the retention of crucial image features.

This method was utilized to resize images while ensuring smooth pixel transitions, minimizing distortions, and preserving the quality of the resized images. This technique helped maintain the spatial structure of the images, ensuring that key features were preserved during the scaling process. The scaling step ensures that the images are prepared for training and validation, enabling the model to process the data effectively and consistently during its development phase.

It is a crucial preprocessing stage in preparing data for machine learning (ML) algorithms, especially for image-based datasets such as CT scans.

By guaranteeing data consistency and ideal scaling for training, this procedure improves the stability and effectiveness of the model. These crucial steps are part of the process:

- **Standardization of Feature Values:** The values of pixels or features are set to a standard range, often **0 to 1**. In addition to ensuring that the model interprets and learns from the input data efficiently, this reduces variability throughout the dataset.
- **Increasing the Consistency of Data:** Brightness, contrast, and other discrepancies are reduced by normalizing feature scales or pixel intensities. This improves the model's ability to recognize significant patterns by producing a dataset that more closely matches its assumptions.
- **Increasing the Effectiveness of Training:** Faster and more stable optimization is encouraged by data scaling. It ensures smoother training and more reliable outcomes by addressing problems such as vanishing or exploding gradients during backpropagation.

An essential step in evaluating a trained machine learning model's performance is testing. This step assesses the model's capacity to accurately predict new, unseen data. The testing approach includes the following elements:

- **Dataset Partitioning:** The dataset is divided into three subsets:
- **Training Data (70%)** – Used to train the model.
- **Validation Data (15%)** – Helps fine-tune parameters and prevent overfitting.
- **Testing Data (15%)** – Used exclusively for evaluating the final model.

This organized method ensures a thorough evaluation of the model across various development stages.

Hold-Out Validation: The hold-out validation method involves training the model on the training set, refining it

***Corresponding Author:** Dr. B. P. Pradeep Kumar

using the validation set, and assessing its performance on the testing set. This approach ensures the model performs well on unseen data while avoiding overfitting.

Evaluation Metrics: The model's performance is gauged using key metrics:

- **Accuracy** – Percentage of correct predictions across all instances.
- **Precision** – Proportion of true positive predictions among all positive predictions.
- **Recall** – Proportion of actual positive cases correctly identified.
- **Area Under the Curve (AUC)** – Evaluates the model's ability to compare between classes effectively.

ResNet-50, a **50-layer Residual Network**, tackles the vanishing gradient issue through the use of layer-bypassing shortcut connections. This architecture allows deeper networks to maintain high accuracy, making it particularly effective in complex image classification tasks, such as identifying lung cancer from CT scan **images**.

ResNet-50 Training Process

1. **Transfer Learning:** The model utilizes pre-trained weights from the ImageNet dataset, harnessing existing feature extraction capabilities to conserve computational resources and enhance performance.
2. **Data Preparation:** CT scan images are adjusted to **224 × 224** pixels and normalized to align with the ImageNet dataset's scale. The images are then organized into batches for efficient processing.
3. **Model Adaptation:** The ResNet-50 architecture is modified for lung cancer detection. The final fully connected layers are replaced with a custom classification layer to identify four categories: **adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal tissue**. A **softmax activation function** is used to generate class probabilities.
4. **Training Parameters:** Hyperparameters are established to ensure effective learning. In this research:
    - Learning rate: **0.001**
    - Batch size: **16**
    - Epochs: **20**

These values are selected through initial testing to achieve a balance between effective training and overfitting prevention.

5. **Optimization and Loss Function:** The **Adam optimizer**, renowned for its capability to handle sparse gradients, is used to update model weights. The **categorical cross-entropy loss function** measures the difference between actual and predicted class labels, guiding the optimization process.
6. **Training Implementation:** During the training phase, the model iteratively minimizes the loss function using the training dataset. The model's performance is assessed on a separate validation set after each epoch, ensuring learning effectiveness and preventing overfitting. Training continues until performance stabilizes.
7. **Regularization Methods:** To mitigate overfitting, techniques such as dropout and weight decay are employed. Dropout involves randomly deactivating certain neurons during training, improving the model's ability to generalize.
8. **Performance Metrics:** The trained model's effectiveness is evaluated using the following metrics:
    - **Accuracy:** Proportion of correct predictions.
    - **Precision:** Ratio of true positives to all predicted positives.
    - **Recall:** Ratio of true positives to all actual positives.
    - **Area Under the Curve (AUC):** Indicates the model's ability to differentiate between classes.

The resulting ResNet-50 model exhibits **enhanced performance** in classifying different categories of lung cancer. By combining pre-trained architectures with fine-tuning, the model achieves accurate predictions while maintaining computational efficiency. The methodology ensures reliable results through iterative optimization, hyperparameter adjustment, and thorough evaluation.

The disease classification stage utilizes machine learning techniques to classify CT scan images into different lung cancer categories. Convolutional Neural Networks (CNNs) are crucial in this procedure, employing extracted features to analyze and categorize images according to the presence and type of lung cancer.

Disease Classification Procedure

- **Feature Extraction:** A CNN extracts high-level features from CT scan images. These features capture unique patterns and structures associated with various lung cancer types, allowing the model to differentiate between healthy tissue, **squamous cell carcinoma, large cell carcinoma, and adenocarcinoma**.
- **Classification Layers:** The extracted features are passed through the CNN's classification layers, typically comprising fully connected layers and a softmax activation function. The model assigns probability scores to each class, with the highest-scoring class selected as the prediction.
- **Model Evaluation:** Classification performance is assessed using metrics such as **accuracy, precision, recall, and Area Under the Curve (AUC)**. These metrics ensure reliable predictions by evaluating the model's ability to distinguish between different lung cancer types.

***Corresponding Author:** Dr. B. P. Pradeep Kumar

- **Decision-Making:** Each CT scan image is assigned a class label based on the model's predictions. The final classification is determined by choosing the class with the highest probability score.

Accuracy: Quantifies the proportion of correctly identified cases out of all instances, offering an overall measure of model performance.

Precision: Represents the ratio of true positive predictions to all positive predictions, emphasizing the importance of minimizing false positives.

Recall: Also recognized as sensitivity, measures the proportion of correctly identified true positives, highlighting the model's ability to detect lung cancer.

F1 Score: Balances precision and recall by calculating their harmonic mean, making it useful in scenarios with imbalanced class distributions.

Area Under the Curve (AUC): Assesses the model's ability to differentiate between positive and negative cases across all classification thresholds, with higher AUC indicating superior performance.

Evaluation of the Validation Set: To track model performance during training, a validation set comprising **15%** of the dataset is used. Routine evaluations guide adjustments to enhance performance.

Assessment of the Testing Set: Following training, a testing set comprising **15%** of the dataset provides an objective assessment of the model's capacity to generalize to previously unseen data.

Confusion Matrix: A confusion matrix provides a thorough understanding of performance and opportunities for improvement by visualizing true positives, true negatives, false positives, and false negatives.

Performance Visualization: Visual aids such as ROC and precision-recall curves are used to depict performance indicators. These graphs help compare models and offer intuitive insights into model performance.

RESULTS

This section evaluates the effectiveness of several machine learning methods, including **CNN, ResNet-50, Inception V3, and Xception**, for classifying lung cancer using a set of CT scan images. To provide a thorough assessment, the evaluation uses measures such as confusion matrices, training and validation accuracy and loss, and comparative analysis.

Confusion Matrix: The confusion matrix is used to evaluate the correctness of the trained models. The CNN model achieved an overall accuracy of **85.4%**. Specifically, **247 benign predictions were correctly classified in 98.9% of cases**.

The accuracy for malignant predictions was **100%**. The model also demonstrated a strong ability to distinguish between normal and malignant tissues, with **72.4%** accuracy in classifying normal instances, as shown in **Figure 3**.

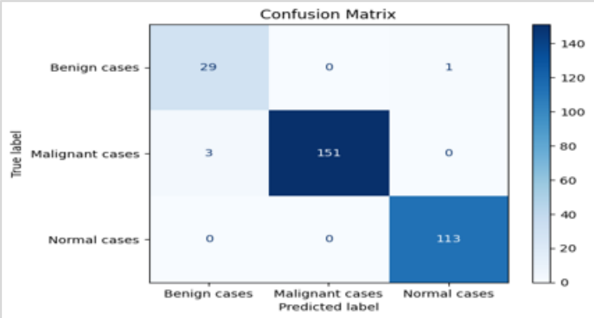

Fig. 3. Confusion Matrix

Validation Accuracy vs. Epochs: Validation accuracy measures the model's ability to generalize to unseen data. In **Figure 4**, the CNN model's validation accuracy remained consistent across epochs, peaking at **91.10%**, indicating robust generalization capability.

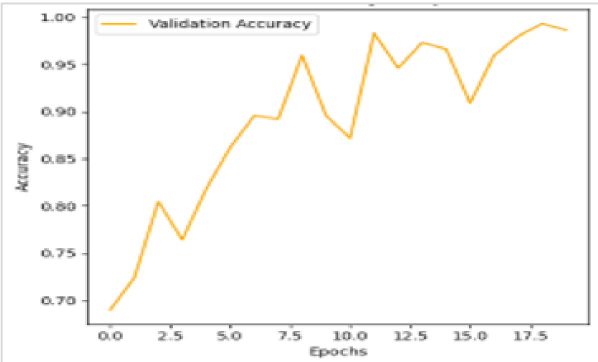

Fig. 4. Validation Accuracy vs. Epochs

Training Loss vs. Epochs: Training loss measures errors during the learning process. The CNN model attained a minimal training loss of **0.002**, as shown in **Figure 5**, signifying a highly effective learning process with few errors.

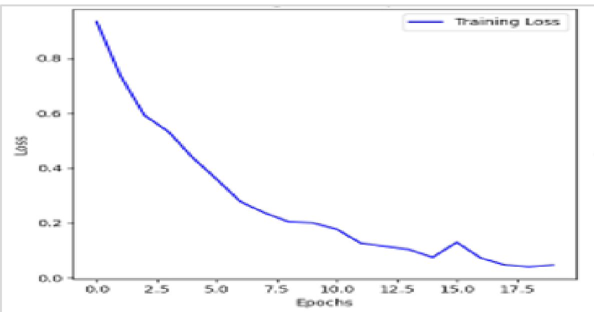

Fig. 5. Training Loss vs. Epochs

*Corresponding Author: Dr. B. P. Pradeep Kumar

Validation Loss vs. Epochs: Validation loss evaluates model performance on unseen data. The CNN model effectively minimized errors, with validation loss steadily decreasing to **0.352**, confirming reliable performance, as shown in **Figure 6**.

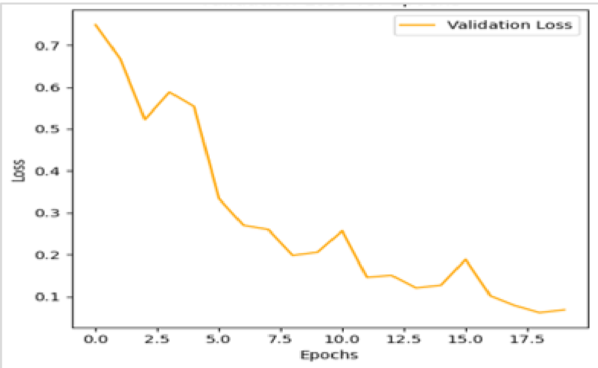

Fig. 6. Validation Loss vs. Epochs

Comparative Analysis: The CNN consistently outperformed ResNet-50, Inception V3, and Xception across several performance metrics. It achieved the lowest testing loss of 0.328, a testing accuracy of 92%, and a testing AUC of 98.21%. These results confirm the superiority of the CNN model for this specific application.

CONCLUSION

The present study demonstrates significant advancements in the classification of lung nodules using Convolutional Neural Networks (CNNs) and CT scan images. The effectiveness of models such as AlexNet and 3D CNNs in differentiating between benign and malignant nodules, thereby improving early lung cancer diagnosis, is demonstrated by their consistently high accuracy, which frequently exceeds 99%. CNNs perform exceptionally well in terms of recall, F1-score, and accuracy, all of which are essential for timely and effective treatment.

However, larger and more diverse datasets are needed to further improve these models. Future research should focus on transfer learning and extending machine learning applications to other types of cancer. Ultimately, CNN-based medical imaging provides a rapid, accurate, and non-invasive diagnostic tool that supports clinical decision-making and improves patient outcomes.

REFERENCES

1. Augustine, R. (2025). *Efficient deep learning architectures for brain tumor diagnosis: A comparative study of Xception, MobileNet, MLUNet and DeepLabS* (SSRN Working Paper No. 5236459). SSRN. https://doi.org/10.2139/ssrn.5236459
2. Augustine, R. (2025). *Machine learning for personalized juvenile diabetes risk stratification: An in-depth comparative study* (SSRN Working Paper No. 5177119). SSRN. https://ssrn.com/abstract=5177119
3. Rangaiah, P. K. B., Huss, F., et al. (2025). Precision diagnosis of burn injuries using imaging and predictive modeling for clinical applications. *Scientific Reports, 15*, Article 7604. https://doi.org/10.1038/s41598-025-92096-4
4. Shanthamallappa, M., & Pradeep Kumar, B. P. (2025). Enhanced perceptual wavelet packet features for spontaneous Kannada sentence recognition under uncontrolled conditions. *International Journal of Speech Technology*. https://doi.org/10.1007/s10772-024-10156-y
5. Naresh, E., Ashwitha, A., Reddy, K. T., & Srinidhi, N. N. (2025). The burn grafting image reclamation redefined with the peak-valley approach. *Critical Reviews in Biomedical Engineering, 53*(2).
6. Kumar, B. P. P., et al. (2022). Advanced electric vehicle charging infrastructure using Internet of Things. In *2022 3rd International Conference on Communication, Computing and Industry 4.0 (C2I4)* (pp. 1–6). IEEE. https://doi.org/10.1109/C2I456876.2022.10051342
7. Naresh, E., Murthy, S. V. N., Pareek, P. K., Reddy, K. T., Shiva Darshan, S. L., et al. (2024). DevOps life cycle implementation on real life scenarios. In *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1269–1271). IEEE. https://doi.org/10.1109/ICKECS61492.2024.10617200
8. V. C., R., S. A., D., B., & V., S. (2023). Speech-to-text transfiguration in language numerals for perpetual deaf patients. In *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICEEICT56924.2023.10157113
9. Ravikumar, J., Shankar, B. B., & Manjunath Kamath, K. (2025). Optimized anti-interference dynamic integral neural network approach for dementia prediction in health care. *Knowledge-Based Systems, 321*, Article 113723. https://doi.org/10.1016/j.knosys.2025.113723
10. Rangaiah, P. K. B., & Augustine. (2025). From microwave measurement to application: Enhancement of fat-intrabody communication by advanced computational techniques. In *2025 16th German Microwave Conference (GeMiC)* (pp. 490–493). IEEE. https://doi.org/10.23919/GeMiC64734.2025.10979071
11. Surendiran, J., Reenadevi, R., Vidhya, R. G., Sivasankari, S. S., & Balaji, N. (2022). IoT-based advanced electric vehicle charging infrastructure. In *2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP)* (pp. 1–6). IEEE. https://doi.org/10.1109/CCIP57447.2022.10058649

***Corresponding Author:** Dr. B. P. Pradeep Kumar

12. Pradeep Kumar, B. P. (2025). A comparison of CNN, RNN, and FNN algorithms to investigate effective diabetes prediction. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-025-10333-5
13. Naresh, E., Hemavathi, P., Padmavathi, S., Srinidhi, N. N., Pradeep Kumar, B. P., Karthik, V., & Mallik, S. (2024). Autonomous garbage accumulation robot using Internet of Things. *Journal of Machine and Computing, 4*(2), 431–440.
14. R., L., R., S. A., Ibrahim, M. M., & V., S. (2023). Hybrid threshold speech enhancement scheme using TEO and wavelet coefficients. In *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICEEICT56924.2023.10156921
15. Kumar, B. P. (2018). Analysis of ASL recognition system for aligned RGB-D image. *Journal of Advanced Research in Dynamical and Control Systems*.
16. Mallanna, S. D. (2018). High gain perfect matched inset-fed rectangular microstrip patch antenna for 2.4 GHz frequency. *Journal of Advanced Research in Dynamical and Control Systems, 10*(13, Special Issue), 46–52.
17. Pramod, K. B., Kumaraswamy, H. V., Prathap, C., & Swamy, M. (2013). Design and analysis of UHF BJT feedback oscillator using linear and non-linear simulation. In *2013 International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA)* (pp. 1–6). IEEE. https://doi.org/10.1109/C2SPCA.2013.6749386
18. Ahmed, M. (2025). IoT-based battery management system in the electric vehicle charging station. In T. A. Nguyen (Ed.), *Digital twin, blockchain, and sensor networks in the healthy and mobile city* (pp. 431–451). Elsevier. https://doi.org/10.1016/B978-0-443-34174-8.00022-3
19. Somyaji, A. D., Surakasi, L., Puranik, A., & Kannan, D. (2025). Melody mixer: Novel music generation by mixing styles with variational autoencoders. In *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)* (pp. 1–5). IEEE. https://doi.org/10.1109/COMP-SIF65618.2025.10969893
20. D. J., B., G. G., & M. N. (2025). Fake job post detection using machine learning. In *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)* (pp. 1–6). IEEE. https://doi.org/10.1109/COMP-SIF65618.2025.10969867
21. Chand, S. R., Ahmed, U., Ismail, T. M., Chand, S. B., & P. C. (2025). Email validator dashboard: A comprehensive tool for accurate email validation. In *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)* (pp. 1–8). IEEE. https://doi.org/10.1109/COMP-SIF65618.2025.10969888
22. Kumar, R. J., Biswas, A., Hafeeza, B. B., & Dhaanya, D. (2025). Emotion classification in text using machine learning and LSTM. In *2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF)* (pp. 1–7). IEEE. https://doi.org/10.1109/COMP-SIF65618.2025.10969881
23. Tabassum, M., Neha, R., & Khanum, A. (2025). ChargeSmart: Effortless EV charging with real-time slot booking. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11035724
24. Purushotham, R., Mahato, R., & Kumar, S. M. V. (2025). AI-powered plagiarism detection: A web-based system for text integrity. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11036044
25. Afnar, R., Raheem, R., Ranjanyou, N., Ravikumar, J., & Veena, S. (2025). Hand motions based virtual AI mouse. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11035082
26. Harshvardhan, M., Devineni, M., & Raju, P. A. (2025). Framework for tuberculosis detection. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11035324
27. Reddy, L. G., Chethana, M., & Likhita, D. (2025). Improving DDoS attack identification and mitigation in SDN through the use of ensemble online machine learning framework. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11035808
28. Khan, R. M., Kabir, S., & Raj, Y. (2025). Driver drowsiness detection system. In *2025 3rd International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICSSES64899.2025.11010064
29. Kousar, F., Namratha, R. N., & Yogesh, D. (2025). Leveraging AI for advancing health and behavioral optimization. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11035259

***Corresponding Author:** Dr. B. P. Pradeep Kumar

30. Hari Gokul, S. S., Gow, A., & Adarsh, L. (2025). Remote monitoring for water level of bridges and flood zones. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11035864
31. John, A., Harshitha, K., & Pooja, D. (2025). Leveraging histopathological images for detection and diagnosis of colonic carcinoma. In *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICKECS65700.2025.11035505.