iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73626

Breast Cancer Classification Using CNN and SVM: A Hybrid Approach

Swathi K.S[1], Dr. Nirmala G[2], Dr. Sujatha S R[3], Dr. Raviram V[4]
[1]*Student,* [2]*Associate Professor,* [3,4]*Professor, Dept. of Computer Science Engineering,Sri Siddhartha Institute Of Technology, Sri Siddhartha Academy of Higher Education.Tumkur, India*

Abstract: Breast cancer remains one of the most prevalent and life-threatening diseases affecting women worldwide. According to the World Health Organization, early detection and accurate diagnosis play a crucial role in reducing mortality rates and improving treatment outcomes. Despite advancements in diagnostic technologies, manual analysis of mammogram images is time-consuming, prone to variability, and requires expert radiological interpretation. As a response to these challenges, this study proposes an innovative and efficient hybrid machine learning framework that combines the deep learning capabilities of Convolutional Neural Networks (CNNs) with the classification strength of Support Vector Machines (SVMs) for breast cancer detection and classification from mammographic images. The model leverages the CBIS-DDSM (Curated Breast Imaging Subset of DDSM) dataset—a well-established benchmark for mammographic image analysis. The images undergo a series of pre-processing steps including greyscale normalization, contrast enhancement, resizing, and noise reduction, all of which aim to ensure consistent quality and effective feature learning. A custom CNN architecture is then designed to extract high-level features from the pre-processed images. This network is optimized for capturing complex patterns such as masses, calcifications, and tissue asymmetries commonly observed in breast cancer cases. Unlike conventional end-to-end CNN classification, this study uses the CNN primarily for deep feature extraction .The extracted features are subsequently passed to an SVM classifier, which constructs a decision boundary to accurately separate benign from malignant cases. This hybrid model addresses several challenges inherent to medical image analysis: it mitigates the risks of over fitting associated with deep learning models trained on limited data and improves classification performance on imbalanced datasets through the SVM's generalization capability. The proposed hybrid CNN-SVM model achieves a classification accuracy of 91.7%, with competitive precision, recall, and F1-scores, highlighting its potential effectiveness in real-world clinical scenarios. This study's contributions are multifold: the development of a novel hybrid classification framework, the successful application of deep learning techniques for mammographic image analysis, and the demonstration of improved diagnostic accuracy through AI-driven methods. The research underscores the importance of interdisciplinary approaches combining medical imaging, artificial intelligence, and statistical learning for advancing cancer diagnostics. In future work, the integration of transfer learning, explainable AI, and real-time decision support systems could further enhance the diagnostic reliability and acceptance of such tools in clinical environments. The findings of this study pave the way for future advancements in computer-aided diagnosis systems and support the global effort to combat breast cancer through technology.
Keywords: CNN, Breast Cancer, SVM, Mammogram Images.

I. INTRODUCTION

Breast cancer is a major global health concern and remains one of the most commonly diagnosed cancers among women. According to the World Health Organization (WHO), over 2.3 million new cases of breast cancer are diagnosed annually, and it continues to be a leading cause of cancer-related mortality. Early detection significantly improves the chances of successful treatment and long-term survival, making timely and accurate diagnosis a critical component in patient care. Mammography is currently the most widely used imaging technique for early detection and screening of breast cancer. It provides non-invasive visualization of breast tissue and enables the identification of abnormalities such as masses and micro calcifications. However, the interpretation of mammographic images is often a complex and subjective task that depends on the expertise of radiologists. Variability in interpretation, dense breast tissue, and subtle pathological signs can result in missed or incorrect diagnoses. To address these limitations, there is a growing demand for automated, accurate, and efficient computer-aided diagnostic (CAD) systems.

Despite the advances in imaging modalities and diagnostic tools, breast cancer detection still faces several challenges. Manual interpretation is time-consuming and subject to human error, especially in resource-limited settings where trained radiologists may not always be available.

Furthermore, the characteristics of breast lesions can vary significantly in terms of size, shape, texture, and intensity, making it difficult to distinguish benign from malignant tumors through visual inspection alone. Another significant issue is the class imbalancein medical datasets, where benign cases often outnumber malignant ones. This imbalance can affect the performance of conventional machine learning algorithms, leading to biased classification and lower sensitivity in detecting malignant cases. Additionally, over fitting is a common concern when deep learning models are trained on relatively small or non-diverse datasets, as is often the case in the medical domain due to privacy and data availability constraints.

Recent years have witnessed an increasing adoption of Artificial Intelligence (AI) and Machine Learning (ML) in healthcare, particularly in medical image analysis. Among these, Convolutional Neural Networks (CNNs) have emerged as a powerful class of deep learning models capable of automatically learning hierarchical features from raw image data. CNNs have demonstrated remarkable success in tasks such as image classification, segmentation, and object detection across various medical imaging modalities including X-rays, CT scans, MRIs, and mammograms. However, while CNNs are excellent at feature learning and representation, they may not always provide optimal classification performance when trained end-to-end on limited datasets. To overcome this, hybrid approaches that combine deep learning with traditional machine learning classifiers have gained attention. One such technique involves using CNNs for feature extraction and Support Vector Machines (SVMs) for final classification.

II. LITERATURE SURVEY

Breast cancer detection, predominantly reliant on mammographic imaging, has witnessed significant advancements driven by the integration of deep learning and artificial intelligence. This section reviews key recent contributions in the field.A hybrid method leveraging deep convolutional networks and belief networks was proposed in [1], employing a structured support vector machine (SSVM) in combination with conditional random fields (CRF) for the classification of mass regions in mammograms. Among the two, the CRF framework demonstrated superior efficiency in both training and inference stages. Additionally, a Full-Resolution Convolutional Network (FrCN) was introduced, evaluated using four-fold cross-validation on the INbreast X-ray mammogram dataset. This model achieved remarkable performance metrics, including an F1-score of 99.24%, accuracy of 95.96%, and a Matthews Correlation Coefficient (MCC) of 98.96%.

Further advancing detection strategies, a novel architecture named BDR-CNN-GCN was proposed in [2], combining an 8-layer convolutional neural network (CNN) with a Graph Convolutional Network (GCN). When tested on the MIAS dataset, it yielded an impressive accuracy of 96.10%, highlighting the benefit of capturing spatial relationships through graph modelling.

An enhanced object detection system using a modified YOLOv5 framework was explored in [3], achieving 96.50% accuracy and 93.50% MCCin the detection and classification of breast tumors. This architecture outperformed earlier iterations such as YOLOv3 and Faster R-CNN, showing the evolution of real-time object detection in the medical domain.To address the challenge of variability in mammographic features, [4] and [5] introduced the Diversified Features-Based Breast Cancer Detection (DFeBCD) method. By combining feature extraction with emotion-inspired learning and an integrated classifier, they achieved a classification accuracy of 80.30%, particularly useful for binary classification between normal and abnormal cases.

Overfitting, a common issue in deep networks, was mitigated in [6] by implementing transfer learning (TL) alongside a deep CNN architecture. The proposed framework was evaluated on multiple benchmark datasets such as INbreast (95.5%), DDSM (97.35%), and BCDR (96.67%), all reflecting high predictive performance. In [7], a hybrid technique combining Lifting Wavelet Transform (LWT) for robust feature extraction with Moth Flame Optimization and Extreme Learning Machine (ELM) for classification achieved 95.70% accuracy on MIAS and 98.06% on DDSM, highlighting the potential of hybrid bio-inspired optimization methods.

A CNN-based model utilizing the Inception-v3 architecture was evaluated in [8], achieving 0.88 sensitivity, 0.87 specificity, and a high Area Under the Curve (AUC) of 0.946, making it well-suited for clinical applications requiring high diagnostic confidence.Building upon pretrained architectures, [9] presented a comprehensive evaluation of eight fine-tuned CNN models using transfer learning for breast cancer classification. Hybrid models involving MobileNet, ResNet50, and AlexNet reported classification accuracies reaching 95.6%, demonstrating their adaptability across datasets. The authors in [10] and [11] conducted a comparative analysis of four leading CNN architectures—VGG19, InceptionV3, ResNet50, and VGG16—trained on 5000 mammographic images and evaluated on a test set of 1007 images. The results highlighted the strengths of deeper architectures like ResNet50 in capturing complex patterns within mammograms. Convolutional neural networks are utilized to perform classification task in the papers [13], [14], [15] and [16] on the various medical images. Despite considerable progress in breast cancer detection, several challenges persist, including accurate tumor localization, high memory consumption, lengthy processing times, and the extensive data requirements of deep learning models.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

To address these limitations, the following section introduces a novel approach aimed at enhancing the identification and classification of breast cancerand de-noising effect for the analysis of biomedical signals in [17], [18] and [19]

Among the various medical imaging modalities, mammography remains a cornerstone of breast cancer screening due to its high sensitivity to micro-calcifications and its ability to detect early pathological changes. It is particularly effective in identifying small calcified clusters that often serve as early indicators of malignancy. Utilizing low-dose X-rays, mammography captures subtle anomalies such as structural distortions, bilateral asymmetry, nodules, tissue density variations, and—most critically—calcifications. In standard screening procedures, each breast is imaged from two angles, making it especially valuable for women aged 40 and above. When suspicious areas are detected, additional diagnostic mammograms are recommended for more focused examination.

Ultrasound imaging, while offering lower resolution and monochrome output compared to mammography, is advantageous for distinguishing between cystic and solid masses. It typically identifies malignant tumors as amorphous structures with indistinct margins. Magnetic Resonance Imaging (MRI), though costlier and more time-intensive, provides high sensitivity and generates detailed cross-sectional views using non-ionizing radiofrequency waves and magnetic fields. The histopathological examination, often regarded as the gold standard, involves the microscopic analysis of tissue samples. It offers definitive phenotypic insights crucial for both diagnosis and treatment planning. Additionally, thermography, an emerging imaging method that detects abnormal heat patterns in breast tissue, shows potential for non-invasive screening, although further validation is necessary to establish its clinical utility. In recent years, Deep Learning (DL) has emerged as a transformative tool in breast cancer diagnostics. DL models are proficient in complex tasks such as detection, classification, and segmentation of breast lesions. Their ability to process high-dimensional, interrelated data has proven advantageous in both diagnostic and prognostic contexts. Techniques such as transfer learning, automated feature extraction, and generative adversarial networks (GANs) have significantly improved performance. Nevertheless, several obstacles remain—most notably, the limited availability of annotated medical datasets and concerns surrounding patient data privacy. As the synergy between deep learning and medical imaging continues to evolve, it offers promising avenues for earlier detection, more accurate diagnosis, and ultimately, better patient outcomes in the fight against breast cancer.

III. PROPOSED METHODOLOGY

The proposed block diagram fig. 1 illustrates the workflow of breast cancer classification using a hybrid CNN-SVM model. The process begins with the acquisition of mammogram images from the Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) dataset. This is a publicly available and widely recognized dataset containing digitized mammograms with annotated regions of interest and corresponding pathology labels, categorized as benign or malignant. The dataset serves as the foundation for model development and evaluation, providing high-resolution images necessary for accurate analysis.

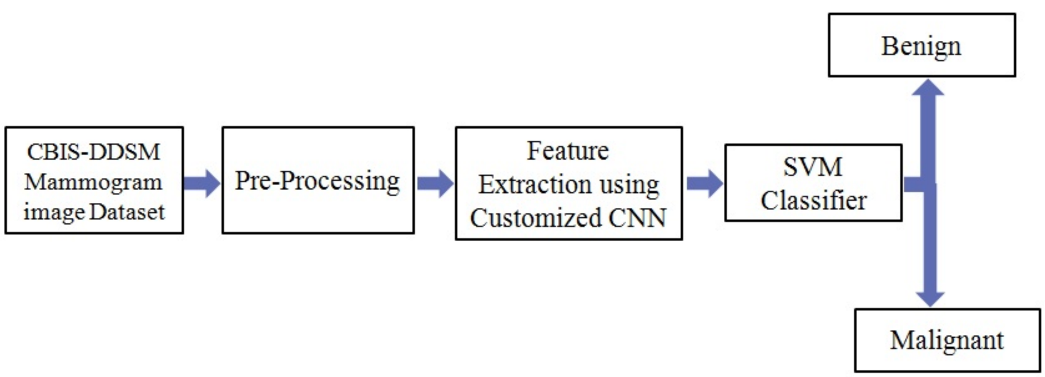

Fig. 1: Proposed block diagram for breast cancer classification

Once the dataset is acquired, the images undergo pre-processing to enhance their quality and improve the performance of the deep learning model. Pre-processing involves several key steps including noise reduction using filtering techniques (such as Gaussian or median filters), image resizing to ensure uniform input dimensions for the CNN, and normalization of pixel intensities to a standard range, which accelerates convergence during training.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

In some cases, region-of-interest (ROI) cropping may be applied based on annotation data, allowing the model to focus on the most relevant parts of the image that contain possible tumor areas. This step is critical in minimizing irrelevant background information and variability, thus ensuring more consistent input data.

The pre-processed images are then fed into a Customized Convolutional Neural Network (CNN) for feature extraction. CNNs are highly effective in image analysis as they automatically learn and extract hierarchical patterns from images. The customized CNN in this architecture is specifically designed to capture intricate patterns and textures present in mammogram images. The network comprises multiple convolution layers that detect low-level and high-level features, pooling layers to reduce dimensionality and computational complexity, and activation functions like ReLU that introduce non-linearity to the model. The final layers of the CNN produce a rich feature vector that encapsulates the most significant characteristics of the input image.

Instead of using the CNN for classification, the output feature vector is passed to a Support Vector Machine (SVM) classifier. SVMs are renowned for their strong classification capability, especially on small to medium-sized datasets, by maximizing the margin between classes. In this system, the SVM acts as a powerful decision boundary classifier that categorizes the mammogram images into two distinct classes: benign or malignant. Benign lesions are non-cancerous and generally do not pose immediate health risks, though they may require periodic monitoring, whereas malignant lesions are cancerous and necessitate prompt medical intervention.

The final stage of the system outputs the prediction result, classifying the input mammogram image as either benign or malignant. This hybrid architecture effectively leverages the feature learning capabilities of the CNN along with the robust classification power of the SVM, providing superior accuracy compared to conventional models. The use of the CBIS-DDSM dataset along with tailored pre-processing and customized CNN design ensures high-quality inputs and reliable outputs, making the system a valuable tool for assisting radiologists in the early detection and diagnosis of breast cancer.

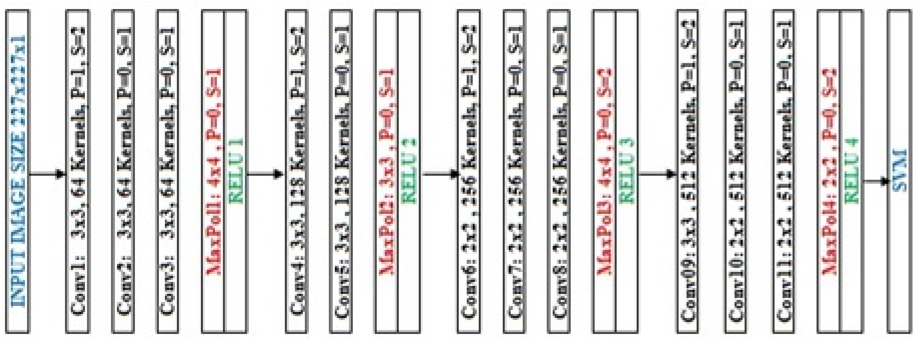

Fig.2: Customized CNN architecture

The feature extraction phase of the proposed model employs Customized Convolutional Neural Network (CNN) architecture, designed specifically to capture the complex texture and structural patterns of mammogram images depicted in fig. 2. The network consists of a total of 11 convolutional layers, which systematically learn hierarchical features from low-level edges and textures to high-level shapes and structures indicative of benign or malignant lesions. The convolutional layers use filter sizes of 3×3 and 2×2, which are optimal for capturing fine-grained details in medical images. The network begins with 64 kernels (filters) in the initial layers and progressively increases the number of filters to 512 kernels in the deeper layers. This gradual increase allows the network to extract increasingly abstract and complex features as the data progresses through the layers. Additionally, to control the spatial resolution and prevent overfitting, four max-pooling layers are integrated at appropriate stages of the network. These max-pooling layers use pooling window sizes of 3×3 and 4×4, combined with suitable padding and stride values to preserve essential information while reducing computational complexity. Padding ensures that the spatial dimensions are maintained where necessary, and strides regulate the step size of the pooling operation. This design enables the CNN to maintain a balance between feature richness and computational efficiency. The deep feature maps generated from the last convolutional layer are then flattened and passed to the Support Vector Machine (SVM) classifier for final classification into benign or malignant categories.

A. Dataset

The proposed breast cancer classification model was evaluated using the Digital Database for Screening Mammography (DDSM), which comprises a large and diverse set of mammogram images. The dataset contains a total of 55,885 images, which are categorized into 48,596 benign (B) and 7,289 malignant (M) images.

To effectively train and assess the performance of the model, the dataset was randomly partitioned into two subsets: 80% for training and 20% for testing. This strategy ensures that the model is trained on a broad and representative sample of the data while reserving a separate portion for unbiased performance evaluation. Specifically, the training set consisted of 44,708 images, of which 38,877 were benign and 5,831 were malignant. The testing set included the remaining 11,177 images, containing 9,719 benign and 1,458 malignant samples. This careful split maintains the natural class distribution of the dataset and provides a realistic scenario for model validation. The large training set enables the deep learning model to learn robust and discriminative features, while the testing set allows for a thorough evaluation of the model's generalization ability on unseen data.

IV. RESULTS AND DISCUSSION

The classification performance of the proposed CNN–SVM hybrid model was evaluated using a validation confusion matrix, which provides a detailed overview of the model's predictive capability across both benign and malignant classes. The matrix summarizes the number of correct and incorrect predictions, enabling the analysis of model strengths and weaknesses in clinical decision-making. In the given matrix, the rows represent the actual classes, while the columns correspond to the predicted classes. Class 0 denotes benign mammogram images, and Class 1 represents malignant mammogram images. Out of a total of 11,177 images in the validation dataset, the model correctly identified 8,951 benign cases (true negatives) and 1,300 malignant cases (true positives). Misclassifications included 768 false positives (benign predicted as malignant) and 158 false negatives (malignant predicted as benign) as shown in fig.3..

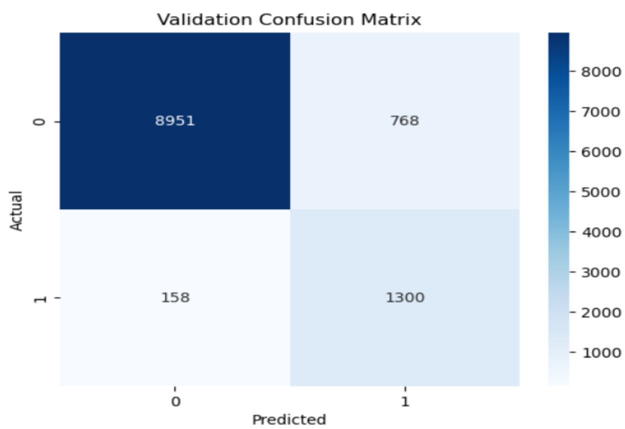

Fig. 3: Confusion matrix

From these results, several important observations can be made. Firstly, the high number of correctly classified benign cases reflects the model's strong ability to differentiate non-cancerous mammograms, which is critical in avoiding unnecessary follow-up procedures for healthy patients. Secondly, the relatively small number of false negatives (158 cases) is clinically significant because missing malignant cases can delay treatment, leading to adverse health outcomes. The recall for the malignant class indicates that the model successfully identified the majority of true cancer cases. This high sensitivity is advantageous in screening scenarios where the primary goal is to ensure that no malignant case is overlooked. However, the precision for malignant detection, reveals that a notable proportion of benign cases were misclassified as malignant, potentially leading to patient anxiety and additional diagnostic procedures.

The overall accuracy of the model was approximately 91.7%, demonstrating its robustness in classifying mammogram images. Furthermore, the F1-score for the malignant classes reflects a balanced trade-off between precision and recall. The relatively higher false positive rate suggests that while the model prioritizes sensitivity, further optimization is needed to enhance precision. Possible improvements could include fine-tuning hyperparameters, incorporating more representative training data for borderline cases, or integrating advanced post-processing techniques to reduce misclassification noise.The proposed CNN–SVM hybrid model achieves this goal effectively, minimizing the likelihood of missing cancerous cases. Although the false positive rate remains a limitation, it is generally preferable in early detection systems to err on the side of caution by flagging suspicious cases for further examination. Overall, the confusion matrix results validate the model's potential for deployment in computer-aided diagnosis systems, where it can assist radiologists in interpreting mammograms with improved efficiency and accuracy.

V. CONCLUSION

The proposed CNN–SVM hybrid model demonstrated strong classification performance for mammogram images, achieving an overall accuracy of 91.7% with high sensitivity for malignant cases. The confusion matrix analysis confirms the model's effectiveness in correctly identifying the majority of cancerous and non-cancerous cases, making it suitable for early breast cancer detection. While the low false-negative rate ensures that most malignant cases are detected, the presence of false positives indicates a need for further optimization to improve precision. Nonetheless, the model shows significant potential for integration into computer-aided diagnosis systems, where it can assist radiologists in enhancing diagnostic accuracy and reducing workload.

REFERENCES

[1] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," Scientific Reports, vol. 9, no. 1, p. 12495, 2019.
[2] H. M. Frazer, A. K. Qin, H. Pan, and P. Brotchie, "Evaluation of deep learning-based artificial intelligence techniques for breast cancer detection on mammograms: Results from a retrospective study using a BreastScreen Victoria dataset," Journal of Medical Imaging and Radiation Oncology, vol. 65, no. 5, pp. 529–537, 2021.
[3] N. Dhungel, G. Carneiro, and A. P. Bradley, "A deep learning approach for the analysis of masses in mammograms with minimal user intervention," Medical Image Analysis, vol. 37, pp. 114–128, 2017.
[4] A. Sahakyan and H. Sarukhanyan, "Segmentation of the breast region in digital mammograms and detection of masses," International Journal of Advanced Computer Science and Applications, vol. 3, no. 2, pp. 87–92, 2012.
[5] M. Abbas, M. Arshad, and H. Rahman, "Detection of breast cancer using neural networks," LC International Journal of STEM, vol. 1, no. 3, pp. 75–88, 2020.
[6] Y. Eroğlu, M. Yildirim, and A. Cinar, "Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR," Computers in Biology and Medicine, vol. 133, p. 104407, 2021.
[7] T. Mahmood, J. Li, Y. Pei, and F. Akhtar, "An automated in-depth feature learning algorithm for breast abnormality prognosis and robust characterization from mammography images using deep transfer learning," Biology, vol. 10, no. 9, p. 859, 2021.
[8] J. Lee and R. M. Nishikawa, "Identifying women with mammographically-occult breast cancer leveraging GAN-simulated mammograms," IEEE Transactions on Medical Imaging, vol. 41, no. 1, pp. 225–236, 2021.
[9] A. Altameem, C. Mahanty, R. C. Poonia, A. K. J. Saudagar, and R. Kumar, "Breast cancer detection in mammography images using deep convolutional neural networks and fuzzy ensemble modeling techniques," Diagnostics, vol. 12, no. 8, p. 1812, 2022.
[10] D. Kaur and Y. Kaur, "Various image segmentation techniques: a review," International Journal of Computer Science and Mobile Computing, vol. 3, no. 5, pp. 809–814, 2014.
[11] Y. Wang, L. Zhang, X. Shu, Y. Feng, Z. Yi, and Q. Lv, "Feature-sensitive deep convolutional neural network for multi-instance breast cancer detection," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 4, pp. 2241–2251, 2021.
[12] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 8, pp. 1309–1321, 2014.
[13] T. M, K. T G, V. K V, N. Siby, S. Devaraj and R. B R, "Leveraging an Optimized Deep Convolutional Neural Network Architecture for the Automated Diagnosis of Pathological Conditions Through Acoustic Analysis of Human Cough Signatures," 2025 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), Tumakuru, India, 2025, pp. 1-7, doi: 10.1109/ICSSES64899.2025.11009841
[14] Murthy, K. T., and M. N. Eshwarappa. "A Novel Person Authentication Technique Using Electrocardiogram (ECG)." J. Electr. Syst 20 (2024): 393-405.
[15] Hareesh K. N. (2024). Comparative Analysis of Brain Tumor Classification Using CT, MRI, and Fusion of CT and MRI Images with GLCM Features. International Journal of Intelligent Systems and Applications in Engineering, 12(21s), 3844, https://ijisae.org/index.php/IJISAE/article/view/6153
[16] T G, K. ., & M N, E. . (2023). ECG Biometric for Human Authentication using Hybrid Method. International Journal on Recent and Innovation Trends in Computing and Communication, 11(7s), 292–299. https://doi.org/10.17762/ijritcc.v11i7s.7002
[17] T. G. Keshavamurthy and M. N. Eshwarappa, "Review paper on denoising of ECG signal," 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICECCT.2017.8117941.
[18] T. G. Keshavamurthy and M. N. Eshwarappa, "ECG signal de-noising using complementary ensemble empirical mode decomposition and Kalman smoother," 2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Tumkur, India, 2017, pp. 120-125, doi: 10.1109/ICATCCT.2017.8389118.
[19] Keshavamurthy, T. G., & Eshwarappa, M. N, " ECG signal de-noising based on adaptivefilters, "International Journal of Innovative Technology and Exploring Engineering(IJITEE), vol.9(1), pp. 5473–5483, November 2019.http://doi.org/10.35940/ijitee.K1601.119119