International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Breast Cancer Histopathological Image Analysis Using deep learning

Vaibhav Santosh Supekar[1], Dr. G. Sunil Kumar[2], Aman Shaikh[3], Harsh Sakhare[4], Bhakti Sinare[5]
[2] Assistant Professor, Department of Computer Engineering, Sinhgad Academy Of Engineering,Pune
[1,3,4,5]Department of Computer Engineering, Sinhgad Academy Of Engineering, Pune

---

Abstract - Breast cancer is a major global health issue and a leading cause of cancer-related deaths among women, making early and accurate diagnosis essential for improving survival rates. Histopathological examination of breast tissue is the gold standard for detection, but this manual process is time-consuming, highly dependent on expert pathologists, and prone to inter-observer variation. To address these limitations, this project aims to develop an automated breast cancer classification system using Deep Learning, specifically Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs). The BreakHis dataset, containing high-resolution histopathological images categorized as benign or malignant, will be used for training and evaluation. The workflow includes image preprocessing, data augmentation (rotation, flipping, zooming), model design, training, validation, and performance assessment using accuracy, precision, recall, and F1-score. A comparative analysis between CNN and SVM models will highlight their effectiveness and potential for real-world medical use. The proposed system is designed as a computer-aided diagnosis (CAD) tool to assist pathologists by reducing workload, improving diagnostic consistency, and accelerating detection. With strong interpretability and reliability, the model aims to deliver high-confidence cancer predictions and can be extended to multi-class classification and other cancer types in the future.

***Key Words*:** Breast Cancer, Histopathological Images, Early Diagnosis, Artificial Intelligence (AI), Deep Learning (DL), journals,BreakHis Dataset,Accuracy,F1-Score

1.INTRODUCTION

Breast cancer is a major global health concern and remains one of the most common cancers among women. According to the World Health Organization (WHO), early detection of breast cancer significantly increases the chances of successful treatment and survival. Traditional methods of diagnosis include physical examinations, mammography, biopsy, and histopathological analysis. Among these, histopathological image analysis is considered the gold standard for accurate diagnosis. However, it relies heavily on the expertise and experience of pathologists, and even minor errors or delays in diagnosis can have life-threatening consequences. Histopathological images are microscopic images of tissue samples obtained through biopsy. These images contain complex patterns that can indicate the presence or absence of cancer.Manually examining these images is not only time-consuming but also subject to variability in interpretation between different experts. To overcome these challenges, Artificial Intelligence (AI) and Deep Learning (DL) techniques have shown tremendous potential in automating and enhancing the accuracy of image-based cancer diagnosis. This project proposes a deep learning-based solution for the classification of breast cancer using Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs). We use the BreakHis (Breast Cancer Histopathological Image Classification) dataset, which includes thousands of microscopic images categorized into benign and malignant tumors. The goal is to develop and compare models that can accurately classify these images and assist doctors in faster, more accurate diagnosis. Deep learning, a subset of machine learning, enables computers to automatically learn from large amounts of data without being explicitly programmed. In medical image analysis, deep learning models particularly Convolutional Neural Networks (CNNs) have become the de facto standard due to their ability to capture and process spatial features in images. These models can extract deep features from histopathological images, learning subtle patterns that are often difficult for the human eye to detect. The BreakHis dataset, used in this project, is a well-known benchmark in the medical imaging community. It contains over 7,000 images of breast tumor tissue collected at different magnifications (40X, 100X, 200X, and 400X). This diversity allows the model to learn representations that are robust across various levels of detail. The dataset is divided into benign and malignant classes, which are further categorized into subtypes. By training a CNN on this dataset, the model can learn to recognize complex visual cues associated with cancerous tissues. Furthermore, integrating Support Vector Machines (SVM) alongside CNNs provides a comparative angle to the project. While CNNs handle automatic feature extraction and classification, SVMs require manual feature engineering but are known for strong generalization in high-dimensional spaces. The project aims to compare these models not only in terms of accuracy but also in computational efficiency, interpretability, and suitability for deployment in clinical settings. One of the major challenges in histopathological image classification is variability in staining, image resolution, and tissue morphology. To address this, preprocessing techniques such as color normalization, resizing, and noise reduction will be implemented before feeding the data into the model. Moreover, data augmentation will help simulate

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

real-world scenarios and improve the model's ability to generalize. This research not only advances automated diagnosis but also holds potential for real-time integration in pathology labs, especially in low-resource settings where access to expert radiologists or pathologists may be limited. With continuous improvements, such AI-based systems can become vital assistants in clinical workflows, ensuring that no case is overlooked due to human fatigue or oversight. In conclusion, this project sets the foundation for building an intelligent, scalable, and reliable diagnostic system that leverages the power of deep learning for breast cancer detection from histopathological images. It also emphasizes the importance of comparative analysis, dataset quality, and clinical applicability in building real-world medical AI tools.

2. Body of Paper

The proposed project, Breast Cancer Detection using Histopathological Images, is designed as an intelligent, AI-driven diagnostic framework that integrates Convolutional Neural Networks (CNN) for automated feature extraction and Support Vector Machine (SVM) for optimized classification. The system focuses on improving the accuracy and efficiency of breast cancer diagnosis by leveraging deep learning on microscopic biopsy images. The methodology emphasizes medical image preprocessing, hybrid model integration, and explainable output generation to support pathologists in early detection and classification of cancerous tissues.

2.1 System Architecture

The proposed architecture for Breast Cancer Detection using Histopathological Images is designed to provide a systematic flow of data from image input to disease classification. The system consists of two main entities the User and the System, which interact through a sequence of processing stages to deliver accurate and interpretable diagnostic results.

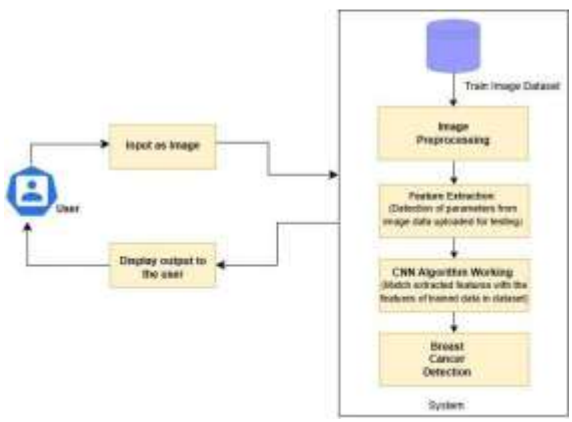

Fig 1-System Architecture

Input as Image: The user uploads a histopathological image sample for analysis through the interface.

Image Preprocessing: The uploaded image undergoes preprocessing operations such as resizing, normalization, noise removal, and color enhancement. This step ensures uniformity and improves model performance.

Feature Extraction: Essential features like cell shape, texture, and color gradients are extracted using advanced image processing and deep learning techniques. This process helps in distinguishing between malignant and benign tissues.

CNN Algorithm Working: The **Convolutional Neural Network (CNN)** analyzes the extracted features and compares them with the trained dataset to identify patterns and similarities. The CNN model learns hierarchical features through convolution and pooling layers.

Classification and Detection: The system performs classification using the trained CNN model or a hybrid CNN-SVM approach, determining whether the image belongs to a **benign** or **malignant** category.

Output Display: The classification result, along with prediction probability and visualization of affected regions, is displayed to the user for interpretation.

2.2 Data Flow Design

The Data Flow Diagram (DFD) represents how data travels and transforms throughout the system. The proposed system consists of three main DFD levels that depict increasing levels of detail.

Level 0 DFD – Context Diagram:At Level 0, the breast cancer detection system is represented as a single, unified process that interacts with two primary external entities — the User and the BreakHis Dataset.The User, typically a medical practitioner or researcher, uploads breast histopathological images to the system for diagnostic analysis. These images serve as the primary input for the system.

The system then retrieves necessary reference samples and data from the BreakHis Dataset, which contains labeled histopathological images categorized as *Benign* and *Malignant*. This dataset supports the system in training the classification model and validating image patterns for accurate prediction.

Once the uploaded image is processed using the trained CNN-SVM model, the system analyzes the tissue features and determines whether the sample shows benign or malignant characteristics.Finally, the system returns the output to the user in the form of a diagnosis result either Benign or Malignant. This level of the DFD emphasizes the overall workflow and high-level interaction between the user and the system, without diving into internal modules or processing details.

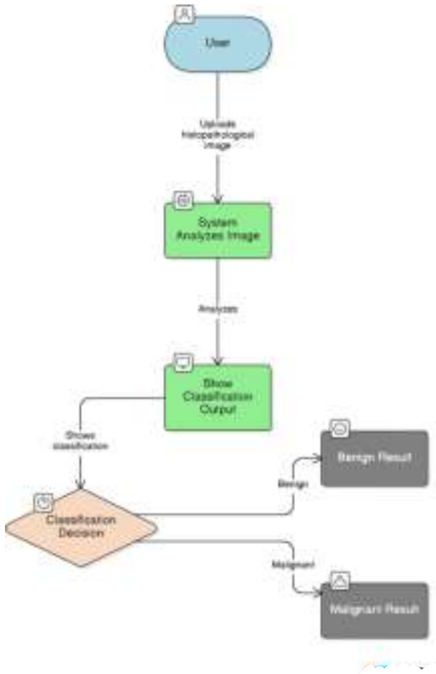

Fig 2- Level 0 DFD-Overall Process Flow

Level 1 DFD – System Process Overview: At Level-1, the single process shown in the context diagram is divided into multiple internal modules to illustrate how the breast cancer detection system functions step-by-step.

The workflow begins with the Image Input and Preprocessing module. In this stage, the system accepts histopathological images uploaded by the user, along with training samples sourced from the BreakHis dataset. Both input sources undergo preprocessing operations such as resizing, normalization, and color correction to ensure uniformity and eliminate noise, making them suitable for model training and testing.

The next phase involves Model Training, where the Convolutional Neural Network (CNN) extracts deep image features, and the Support Vector Machine (SVM) classifier learns to distinguish between benign and malignant tissue characteristics. This stage builds the core predictive capability of the system by learning patterns from the augmented dataset.

Once the model is trained, the system enters the Evaluation and Prediction module. During this process, new or unseen images are passed through the trained CNN-SVM pipeline to identify cancer type. The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure reliability.

The final stage is Result Visualization, where outcomes are presented to the user in a clear and meaningful format. The system displays the predicted label (Benign or Malignant) along with performance metrics, helping users understand both the diagnosis and the efficiency of the model.

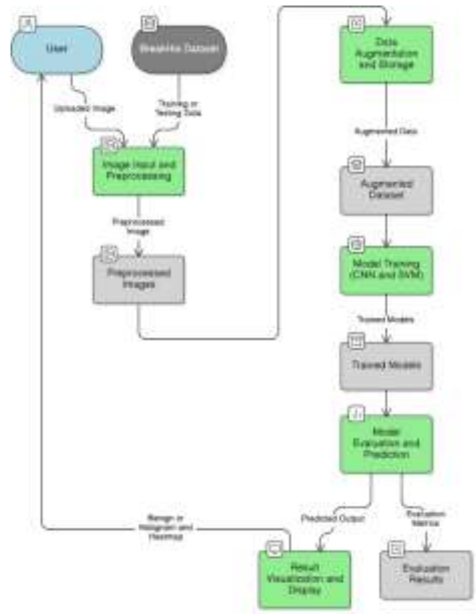

Fig. 3- Level 1 DFD – Data Flow through Processing Stages

Level 2 DFD – Detailed Functional Flow: At Level 2, the system's internal operations are illustrated at a deeper level, showing how each component contributes to the breast cancer diagnosis pipeline.

The process begins within the Preprocessing Module, where raw histopathological images are received from the user or dataset. This unit performs key enhancements such as noise filtering, normalization, and resizing. These steps ensure that variations in lighting, resolution, and artifacts are minimized, providing clean and consistent input for further processing.

Next, the enhanced images move into the CNN Training and Feature Extraction Module. Here, the Convolutional Neural Network learns visual patterns such as cell structures, nuclei density, and tissue irregularities. The CNN transforms these visual details into structured numerical feature vectors that represent essential characteristics of the image. This conversion allows the system to focus on medically relevant patterns instead of raw pixels.

These extracted feature vectors are then passed to the SVM Training and Classification Unit. The Support Vector Machine uses these features to learn the boundary between benign and malignant tissue samples. During prediction, SVM applies this learned decision boundary to classify new images based on their extracted features.

Following classification, the Model Evaluation Unit measures performance using metrics such as accuracy, precision, recall, and F1-score. These metrics provide a reliable understanding of how well the model distinguishes between cancerous and non-cancerous tissue.

To increase transparency, the system incorporates an Interpretability and Visualization Module (Grad-CAM). This component generates heatmaps that highlight the most critical regions in the histopathology image used by the network for prediction. By showing attention regions, the model assists users, especially medical professionals, in understanding the basis of its diagnostic decision.

Finally, the system moves to the Output and Result Display Module, where the user receives a clear diagnostic result *benign* or *malignant* along with supporting visuals and performance measures. If suspicious regions are detected, they are highlighted on the image for improved interpretability and clinical insight.

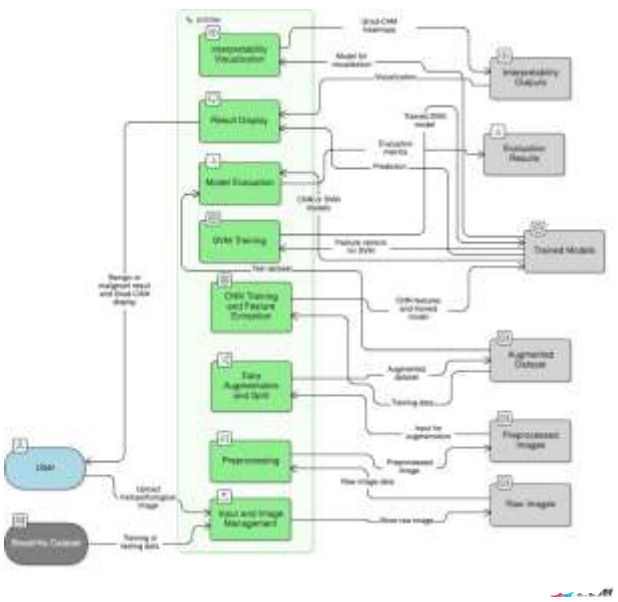

Fig. 4- Level 2 DFD – Detailed Internal Data Flow

Deep Convolutional Neural Network (DCNN) Based Approach:

Designing an efficient CNN architecture is a challenging task, and recent innovations can be broadly categorized into seven areas: spatial feature extraction, depth enhancement, multi-path structures, width expansion, feature-map utilization, channel boosting, and attention mechanisms. CNN models are often considered black-box systems, but they provide vast opportunities for layer-wise customization and optimization.

Several studies have demonstrated strong performance of CNNs for histopathological breast cancer classification. A custom CNN model achieved around 78% accuracy with an error rate near 22%. Transfer learning-based ensemble models have shown significant improvement, reaching approximately 95% accuracy. Pre-trained CNN architectures, particularly those fine-tuned layer-wise, have proven highly effective in medical image analysis tasks. Using data augmentation techniques has also shown to boost CNN performance, achieving close to 94.5% accuracy.

DenseNet-based models have been particularly successful due to their dense residual connections and efficient feature reuse, reporting accuracies above 97% on 200X BreakHis datasets. Other advanced deep learning frameworks, including Deep Belief Networks and capsule-network-based fusion models, have reported performance ranging from around 86% to nearly 94%. Popular architectures such as AlexNet, GoogLeNet, VGG, and ResNet have achieved accuracy levels ranging between 84% and 94% on BreakHis images. Binary classification studies using ResNet and DenseNet have shown accuracy between 91% and 97%, while DenseNet combined with GAN techniques has achieved over 99% accuracy.Overall, DenseNet models consistently demonstrate superior performance due to their advanced feature extraction capabilities, weight-sharing, and faster convergence, making them one of the most effective choices for breast cancer histopathology image classification.

DenseNet121 Base Model:

Consider a training set represented as X={x1,x2,...,xn}X = \{x_1, x_2, ..., x_n\}X={x1,x2,...,xn} where each image sample has ddd features, and the corresponding label set as Z={z1,z2,...,zn}Z = \{z_1, z_2, ..., z_n\}Z={z1,z2,...,zn}. For binary classification, each label vector zkz_kzk contains two elements indicating whether the image belongs to class 1 or class 2. Let wiw_iwi denote the weight matrix associated with the convolutional filters and bib_ibi denote the bias term. The activation function used in the network is represented as ρρρ.

The feature extraction process is carried out through convolutional operations followed by a non-linear activation function. For the kkk-th input image, the feature map at the iii-th convolutional layer is computed as:

Xi=ρ(wi⨀xk+bi)X_i = ρ(w_i ⨀x_k + b_i)Xi=ρ(wi⨀xk+bi)

where ⨀⨀⨀ represents element-wise multiplication. In a deep CNN with fff convolutional layers, the final layer output XfX_fXf is obtained by aggregating the outputs of all previous layers. DenseNet121 achieves this using dense connectivity, where each layer receives inputs from all preceding layers, enabling better feature reuse and stronger gradient flow.

Each dense block increases the number of feature maps based on a growth rate, and bottleneck layers (1×1 convolutions) are used before 3×3 convolutions to reduce computation. Transition layers are introduced between dense blocks to control network size and maintain efficiency.

The model uses Global Average Pooling (GAP) before classification, which reduces spatial dimensions and acts as a regularizer. Dropout is also applied to reduce overfitting and improve generalization performance, making DenseNet121 highly effective for medical image classification.

Datasets: For evaluating the proposed model, the widely recognized and publicly available BreakHis dataset has been used. This dataset is specifically designed for breast cancer histopathology image classification and is frequently adopted in research due to its diversity and reliability. Along with BreakHis, the BACH-2018 dataset and a combined (mixed) dataset were also considered. The mixed dataset consists of images from both BreakHis and BACH-2018, allowing broader performance assessment across different data sources.

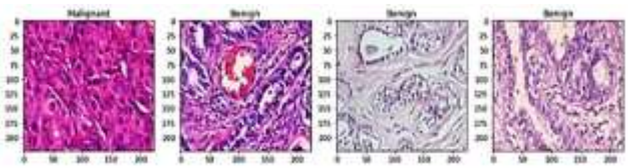

Fig 5-Images From The BreakHis Dataset

These datasets contain images categorized into benign and malignant classes. To ensure fair and consistent evaluation, the data was divided into three parts: 70% for training, 20% for testing, and 10% for validation. This split helps in learning the model parameters, tuning hyperparameters, and measuring final performance effectively.

Before feeding the images into the model, labeling was performed for all samples, followed by data augmentation techniques to enhance generalization. Augmentation included a zoom range of 0.2, horizontal flipping, vertical flipping, and rotation up to 90 degrees. This process increases data diversity and helps the model handle variations in real-world histopathological images.

3. CONCLUSIONS

This survey highlights that CNN and SVM-based methods are powerful tools for breast cancer detection from histopathological images. Hybrid CNN–SVM architectures consistently provide improved accuracy and robustness. Future work should emphasize explainability, dataset expansion, and real-time implementation to support pathologists and enhance clinical decision-making.

ACKNOWLEDGEMENT

We would like to take this opportunity to thank all the people who were part of this seminar in numerous ways, people who gave un-ending support right from the initial stage.In particular we wish to thank Prof. Dr. G. Sunil Kumar as internal project guide who gave their co-operation timely and precious guidance without which this project would not have been a success. We thank them for reviewing the entire project with painstaking efforts and more of his unbanning ability to spot the mistakes. We would like to thank our H.O.D Prof.S.N.Shelke for his continuous encouragement, support and guidance at each and every stage of the project.And last but not the least we would like to thank all my friends who were associated with me and helped me in preparing my project. The project named "Breast Cancer Histopathological Image Analysis Using Deep Learning" would not have been possible without the extensive support of people who were directly or indirectly involved in its successful execution.

REFERENCES

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, ''Cancer statistics, 2021,'' CA, Cancer J. Clin., vol. 71, no. 1, pp. 7–33, 2021.
2. Annual Report, WHO, Geneva, Switzerland, 2021.
3. Breast Cancer Facts and Figures, NBCC, New Delhi, India, Jan. 2022.
4. X.Zhou,C.Li,M.M.Rahaman,Y.Yao,S.Ai,C.Sun,Q.Wang,Y.Zhang, M.Li,X.Li,T.Jiang,D.Xue,S.Qi,andY.Teng,''Acomprehensiverevie w for breast histopathology image analysis using classical and deep neural networks,'' IEEE Access, vol. 8, pp. 90931–90956, 2020.
5. J. de Matos, A. de Souza Britto Jr., L. E. S. Oliveira, and A. L. Koerich, ''Histopathologic image processing: A review,'' 2019, arXiv:1904.07900.
6. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 4700–4708.
7. Y. Zhang, B. Zhang, F. Coenen, and W. Lu, ''Breast cancer diagnosis from biopsy imageswithhighlyreliable randomsubspaceclassifierensembles,'' Mach. Vis. Appl., vol. 24, no. 7, pp. 1405–1420, Oct. 2013.
8. Y. Yari, T. V. Nguyen, and H. T. Nguyen, ''Deep learning applied for histological diagnosis of breast cancer,'' IEEE Access, vol. 8, pp. 162432–162448, 2020.
9. M.Z.Alom,T.M.Taha,C.Yakopcic,S.Westberg,P.Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. S. Awwal, and V. K. Asari, ''A state of-the-art survey on deep learning theory and architectures,'' Electronics, vol. 8, no. 3, p. 292, Mar. 2019.

10. A.Khan,A.Sohail,U.Zahoora,andA.S.Qureshi,''Asurvey ofthe recent architectures of deep convolutional neural networks,'' Artif. Intell. Rev., vol. 53, no. 8, pp. 5455–5516, Dec. 202.
11. N. Wu et al., ''Deep neural networks improve radiologists' performance in breast cancer screening,'' IEEE Trans. Med. Imag., vol. 39, no. 4, pp. 1184–1194, Apr. 2019.
12. V. Rachapudi and G. L. Devi, ''Improved convolutional neural network based histopathological image classification,'' J. Evol. Intell., vol. 14, pp. 1337–1343, Feb. 2020.
13. N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, ''Convolutional neural networks for medical image analysis: Full training or fine tuning?'' IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1299–1312, May 2016.
14. G. Murtaza, L. Shuib, A. W. A. Wahab, G. Mujtaba, H. F. Nweke, M. A. Al-garadi, F. Zulfiqar, G. Raza, and N. A. Azmi, ''Deep learning based breast cancer classification through medical imaging modalities: State of the art and research challenges,'' J. Artif. Intell. Rev., vol. 53, pp. 1655–1720, Mar. 2020.
15. D. R. Sarvamangala and R. V. Kulkarni, ''Convolutional neural networks in medical image understanding: A survey,'' Evol. Intell., vol. 15, no. 1, pp. 1–22, Mar. 2022.
16. M. Masud, A. E. E. Rashed, and M. S. Hossain, ''Convolutional neural network-based models for diagnosis of breast cancer,'' J. Neural Comput. Appl., vol. 34, pp. 11383–11394, Jul. 2022.
17. I. Lahsaini, M. E. H. Daho, and M. A. Chikh, ''Deep transfer learning based classification model for COVID-19 using chest CT-scans,'' Pattern Recognit. Lett., vol. 152, pp. 122–128, Dec. 2021.
18. D. Bardou, K. Zhang, and S. M. Ahmad, ''Classification of breast cancer based on histology images using convolutional neural networks,'' IEEE Access, vol. 6, pp. 24680–24693, 2018.
19. A. Rasheed, M. S. Younis, J. Qadir, and M. Bilal, ''Use of trans fer learning and wavelet transform for breast cancer detection,'' 2021, arXiv:2103.03602.
20. H. Lei, S. Liu, A. Elazab, X. Gong, and B. Lei, ''Attention-guided multi-branch convolutional neural network for mitosis detection from histopathological images,'' IEEE J. Biomed. Health Informat., vol. 25, no. 2, pp. 358–370, Feb. 2021.
21. Y. Feng, L. Zhang, and J. Mo, ''Deep manifold preserving autoencoder for classifying breast cancer histopathological images,'' IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 17, no. 1, pp. 91–101, Jan. 2020.
22. K. George, S. Faziludeen, P. Sankaran, and P. Joseph K, ''Breast can cer detection from biopsy images using nucleus guided transfer learn ing and belief based fusion,'' Comput. Biol. Med., vol. 124, Sep. 2020, Art. no. 103954.
23. Z.Hameed,S.Zahia,B.Garcia-Zapirain, J. J. Aguirre, and A. M. Vanegas, ''Breast cancer histopathology image classification using an ensemble of deep learning models,'' Sensors, vol. 20, no. 16, p. 4373, Aug. 2020.
24. M. Lin, Q. Chen, and S. Yan, ''Network in network,'' 2013, arXiv:1312.4400.
25. S.Ioffe and C. Szegedy, ''Batch normalization: Accelerating deep network training by reducing internal covariate shift,'' in Proc. Int. Conf. Mach. Learn., 2015, pp. 448–456.
26. V. Thakkar, S. Tewary, and C. Chakraborty, ''Batch normalization in con volutional neural networks—A comparative study with CIFAR-10 data,'' in Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. (EAIT), Jan. 2018, pp. 1–5.
27. Y. Mae, W. Kumagai, and T. Kanamori, ''Uncertainty propagation for dropout-based Bayesian neural networks,'' Neural Netw., vol. 144, pp. 394–406, Dec. 2021.