

## Research Article

# A Novel Hybrid Convolutional Neural Network Approach for the Stomach Intestinal Early Detection Cancer Subtype Classification

Md Ezaz Ahmed 

*College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia*

Correspondence should be addressed to Md Ezaz Ahmed; [m.ezaz@seu.edu.sa](mailto:m.ezaz@seu.edu.sa)

Received 7 May 2022; Revised 5 June 2022; Accepted 10 June 2022; Published 24 June 2022

Academic Editor: Arpit Bhardwaj

Copyright © 2022 Md Ezaz Ahmed. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There may be different types of cancer that cause fatal effects in the human body. In general, cancer is nothing but the unnatural growth of blood cells in different parts of the body and is named accordingly. It may be skin cancer, breast cancer, uterus cancer, intestinal cancer, stomach cancer, etc. However, every type of cancer consists of unwanted blood cells which cause issues in the body starting from the minor to death. Cancer cells have the common features in them, and these common features we have used in our work for the processing. Cancer has a significant death rate; however, it is frequently curable with simple surgery if detected in its early stages. A quick and correct diagnosis may be extremely beneficial to both doctors and patients. In several medical domains, the latest deep-learning-based model's performance is comparable to or even exceeds that of human specialists. We have proposed a novel methodology based on a convolutional neural network that may be used for almost all types of cancer detection. We have collected different datasets of different types of common cancer from different sources and used 90% of the sample data for the training purpose, then we reduced it by 10%, and an 80% image set was used for the validation of the model. After that for testing purposes, we fed a sample dataset and obtain the results. The final output clearly shows that the proposed model outperforms the previous model when we compared our methodology with the existing work.

## 1. Introduction

Cancer is a type of body unwanted blood cells that is considered to be the second most frequent occurring disease globally [1–3]. A virus known as the human papillomavirus (HPV) is the cause of body cancer. The virus has the ability to harm cells in the body. The dysplasia cells resemble cancer cells, yet they are not cancerous. Common Intraepithelial Neoplasia (CIN) is the name given to these cells. CIN is divided into three classes based on severity viz. CIN1, CIN2, and CIN3. CIN1 will relapse over time due to the body's immunological system. As a result, the primary objective is to identify CIN2 and CIN3. If the proper sort of body is recognized early on, the treatment will be shown beneficial. Early identification of cancer can boost the five-year survival rate to 91 percent, according to Cancer Statistics [4–6]. Cancer screening procedures that are routinely utilised include the Pap test, HPV testing, colposcopy, and digital angiography. These are useful screening procedures; however, they have limited sensitivity

in identifying CIN2/3+. Furthermore, these examinations necessitate the use of skilled staff and a laboratory setting. Because the entire abovementioned screening test includes human intervention, it is more vulnerable to human mistakes [7–10]. Deep learning and computer vision have been shown to be useful in the healthcare area for medical image categorization. The experimentation in this paper is based on a picture dataset that is accessible through some public open data sources. In June 2017, Kaggle issued a challenge by making thousands of body datasets available to the public dataset via the Intel and Mobile ODT competition. The assignment involved categorising one's body image into one of three categories. We have suggested a novel algorithm to solve the aforementioned problem [11–15].

There may be different types of cancer present in the body. Cancer is the unwanted occurrence of blood cells in different parts of the body [16–25]. According to the body parts, the blood cells grow unnaturally in categories such as skin cancer, intestine cancer, and uterus cancer. In women, it is a kind of

cancer that affects the lower region of the uterus. It is connected to the uterus and the vaginal divider [26–30]. The body resembles a little doughnut with a small aperture in the centre. The posture of a woman’s body changes throughout her life. The body is seen using a tool known as a speculum. By inserting the speculum into the vaginal canal, professional practitioners examine the body and make a prediction [31]. It is caused by a virus called human papillomavirus (HPV) [32–40]. This HPV is found in everyone. Some HPVs are not dangerous, while others might cause abnormal cell development in the body [41–44]. HPV-related illnesses spread gradually. It takes between 10 and 20 years to progress from the precancerous to the cancerous stage. Similarly, cervical malignant development can be avoided if a cervical illness is detected and treated in the premalignant growth stage. With early detection, the rate of survival can be enhanced [45–50]. As indicated, body cancer is broadly categorised into three types: Types I, Type II, and Type III which are the first stage, the second stage, and the last third stage [51–57], respectively. There is a fine boundary between these three types, making it difficult for healthcare personnel to diagnose. The work may get simpler as technology advances. In our work, we have gathered the dataset from different sources and considered our work for the common cancer cell diagnosis.

## 2. Related Work

The craniocaudal (CC) image and the mediolateral oblique (MLO) view of the intestines are supplied simultaneously during the analysis and diagnosis of any type of cancer screening. Mammograms obtained from many perspectives provide more information and characteristics about the lesion than a single scan [1–6]. As a result, mammography classification analysis approaches that are not confined to a single view but rather multiview are beginning to emerge. There are two categorization modes for multiview mammograms [7–10]. Binary segmentation masks or regions-of-interest (ROI) are required to engage in training in one modality. Authors in [11, 12] created a stomach classification method that combines two views of the same intestine and their respective lesion masks, in which each image and associated masks [13, 14] were trained separately, and then a final CNN classifier was trained with characteristics learned from both views. In [15], the authors have built an automatic analytic approach for multiview mammography categorization on this foundation. Using two views and their corresponding masks as input, authors in [16–19] have created a residual neural network to determine if a tumour is malignant or not. In [20], integrated ROIs from four images of both breast mammograms have been used for three-stage classification: abnormal or normal, calcification or tumour, and lastly benign or malignant. Papers [21–27] have suggested a model for obtaining multiscale classification features. The approaches described above still necessitate the use of extra masks or ROI annotations, which necessitate the expenditure of significant resources by a well-trained clinician. In the other mode, approaches for mammography classification analysis that do not require pixel-level or bounding box annotations are

created. In [28], the authors have used four standard views of the right and left breasts to classify tests as Imaging Reporting and Data System (BI-RADS) 1, BI-RADS 2, and BI-RADS 3. Authors in [29, 30] have trained a CNN to distinguish between malignant and nonmalignant tumours using four standard views and heatmaps. In paper [31], authors have established a multiscale approach for cancer calcification classification that pulls information from each mammography image (CC and MLO views). For the multiview challenge, in research work [32–35], they created a convolutional neural network that joined four mammography pictures at the same time. Authors in [36] have introduced a two-stage approach for classifying two-view mammograms. A two-stage convolutional neural network model was used by the tools to classify benign and malignant mass cases. Based on the actual diagnosis process of the doctors combining the CC view and MLO view mammograms, the good performance brought by multiview methods benefits from complementary information between different views. On this basis, we consider both global and local features of mammograms to further improve the expressiveness of the model. An end-to-end neural network model that combines the CC and MLO views is trained only using image-level labels for the whole mammogram benign or malignant classification. Our model can effectively conduct global and local features in spatial dimensions to obtain a relatively more competitive result [37–40]. Many time series models still exploit simple regressive algorithms instead of deep learning [40–43]. Some reasons for this are interpretability, limited data, and low training cost. Attention mechanisms provide a compelling argument, and the results can be used to interpret the reasons for the performance improvements. In the healthcare domain, authors in [44, 45] have presented an interpretable bidirectional recurrent neural network-based approach (HAN-ECG) for the detection of Atrial Fibrillation (AF) based on the ECG recordings. It is a hierarchical attention network that provides three attention modules to realize multiresolution analysis in ECG leading to AF. One of the first papers [46] proposes to utilize 1-D convolution- and self-attention-based Simply Attend and Diagnose (SAnD) architecture with both single-task and multitask strategies for medical multivariate time series data. Since the self-attention mechanism examines relations for all pairs of timestamps, one of its biggest difficulties is the consideration of a long time series. To overcome this, a masking mechanism was used to hide timestamps that are too far in the past and the dense interpolation was applied instead of the addition and layer normalization components after the self-attention module. The authors in [47–53] have developed a novel temporal attention encoder-decoder model, called MTSMFF, for multivariate time series. The authors in [58] have used genetic algorithms for detection and prediction of cancer and critically analysed the state-of-the-art techniques used for cancer research. The authors in [59] have implemented a model with machine learning techniques such as support vector machine, logistic regression, and  $K$ -nearest neighbor (KNN) in breast cancer classification.

### 3. Proposed Methodology

To judge the benign or malignant cancer cells by the whole mammogram or other radio processes, the multiview multifeatures convolutional neural model extracts feature information from the data input given to the convolutional model. The proposed network first extracts features separately through parallel convolution layers and fuses features, then obtains global context information from fused features through the self-attention mechanism, and finally refines important local features through multiplex convolutions. Figure 1 represents the multifeatures extraction model architecture which mainly includes the following:

- (1) Pretrained ResNeXt: a few layers of the pretrained ResNeXt-101 with a  $32 \times 8d$  template are contributed to extracting whole blood cell features.
- (2) Transformer encoder: it extracts features to construct global features to overcome the limitation of the receptive field.
- (3) Multiplex convolutions: they send more refined local features to the classifier to get classification.
- (4) Classifier: it contains a fully connected layer and a softmax layer.

Figure 2 shows the processing of the sample data. We have shown some cancer blood cells of different types of cancer occurring in the body.

Different types of convolutions have different feature extraction capabilities. We propose a multiplex convolutions module to elaborate local features. Different from the standard convolution, the deformable convolution has an offset setting that allows the sampling grid to be deformed freely. The dilated convolution [47] is applied to convolution regions by a filter with a dilation factor. The two convolutions have powerful capabilities to extract features of multiple receptive fields from images while increasing the receptive field without subjoining the number of connections and calculation parameters. Figure 3 shows the processing of the input images through the proposed model which passes through different stages to extract the multifeatures of blood cells and then a different dataset is processed for the validation of the model. Every type of cancer has a common lump of blood and distinguishes it from different body parts. To take full advantage of the characteristics, we gave proposed different layer convolutions, and we design three branches based on standard convolution, deformable convolution, and dilated convolution, respectively. We restore the sequence data from transformer layers into a standard feature map, which is then used as the input for each branch. In addition, global average pooling is concatenated after different convolutions to reduce the number of parameters and overfitting.

Figure 4 shows the testing of the model. The testing images go through different testing phases and give the output in terms of the classification of cancer cells, representing positive or negative.

The proposed method algorithm is shown as follows, and Table 1 represents the different feature cells extracted:

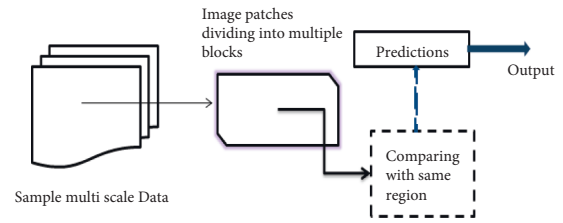


FIGURE 1: Process sampling dataset for multiscale module.



FIGURE 2: Representation of four subtype cancer cell images.

Table 1 gives a description of feature extraction with different types of filters.

ResNet introduces shortcut connections to form residual learning and used the gradient disappearance neural network method to solve the existing problem. Based on ResNet, ResNeXt which is a structural paradigm of split-transform merge replaces the standard convolution with grouped convolution, splits and merges the channel. Considering its advantages that improve the accuracy without increasing the complexity of the parameters and reducing the number of hyper parameters, we choose ResNeXt (ResNeXt with 101 layers and grouped convolution with 32 groups) as a base module to extract spatial features from the whole cancer cells. To save computational resources and reduce the number of parameters, instead of using the integrated ResNeXt-101  $32 \times 8d$  structure, we only use the first  $7 \times 7$  265 convolution, max-pooling, and the first three convolution blocks as the first module of our model. Transfer learning reuses the weights of the pretrained model as the starting point for training our model, making the model robust and easier to train. Based on the idea of transfer learning, we choose the pretrained ResNeXt  $32 \times 8d$  to initialize our ResNeXt module. The pretrained model was first trained on the Instagram dataset that contains  $9.4 \times 10^8$  images and 1,500 labels with weakly supervised learning and then fine-tuned on ImageNet [42]. The multiview and multifeatures extraction of cancer are denoted by Xcc and Xmlo. For each view, we apply the pretrained ResNeXt module for spatial feature extraction from the whole dataset and obtain a 1024-dimension hidden representation vector,

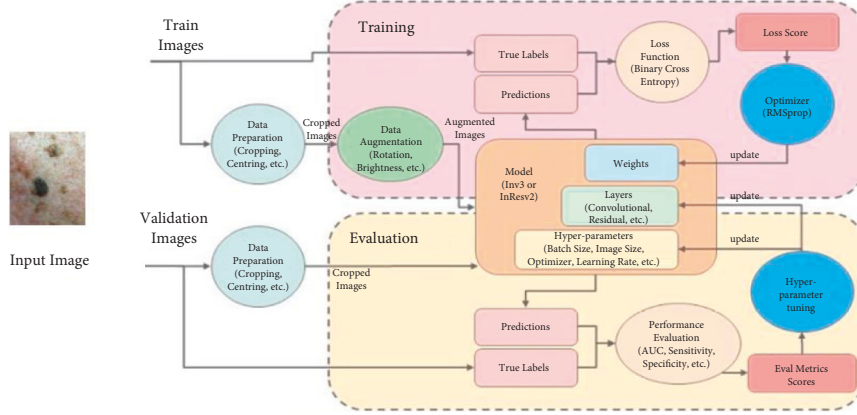


FIGURE 3: Input image undergoing training and validation process.

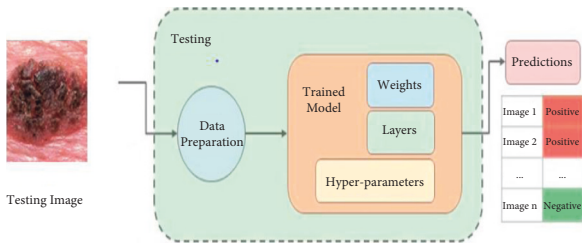


FIGURE 4: Testing process for the outcomes.

TABLE 1: Feature extraction.

	Types of filters	Description
Feature extractor	Canny	Used to detect and extract the features of the edges
	CR	Used to find the $L_1$ regularized parameter value
	VLAD	Used to find the $k$ -dimensional and $L_2$ regularized parameter value

respectively. In the proposed work, we have clearly trained the dataset and then processed it for the validation purpose from the CNN model; finally, the model is tested and obtains the output results. The proposed model outperforms the existing models.

#### 4. Experimental Results

The model includes two subnetworks: a multiview CNN and a multidilated CNN, with the multiview CNN extracting features from multifeatures and feeding them into a multidilated CNN for classification. DLA-SE-Res2NeXt-60 [31] is a multiscale attention network model that introduced multiscale capability, the squeeze-and-excite method, and deep layer aggregation that harvests block- and layer-wise characteristics. MVNN [33] primarily combines a multiscale convolution module with an attention module for feature extraction.

The training set is made up of 90% of the samples in our dataset, whereas the testing set is made up of 35%. Only the training set, which is utilised to optimise our model

parameters, is supplemented. Gradient descent with a stochastic component (SGD), the optimizer is used, with a momentum of 0.9 and a weight decay of 5.1%. A minibatch is defined as a group of eight sample data. The training epochs have been set to 100. The network with a factor of ten reductions in learning rate with an initial value of 0.001, 10 is added every 20 epochs. During the set, the number of transformer layers  $L = 12$  in the transformer encoder module. We use Test Time during inference. The various mathematical parameters used are shown as follows:

$$T = \{(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_n, Y_n)\}. \quad (1)$$

The following function shows the fitting curve function where  $W$  represents the weights of the layers and  $b$  is the bias factor introduced:

$$F(X) = W^T x + b. \quad (2)$$

The loss function is given by

$$L(X, Y, F) = Y - F(X). \quad (3)$$

The various mathematical factors used in the work are calculated using the standard formulas. Some of the mathematical equations are represented as follows:

$$AP = \sum_{k=1}^n \text{precision}(k) \times \Delta \text{recall}(k),$$

$$\text{Precision} = \frac{Ntp}{Ntp + Nfp}, \quad (4)$$

$$\text{Recall} = \frac{Ntp}{Ntp + Nfn},$$

where  $Ntp$  is the number of correctly detected samples,  $Nfp$  is the number of falsely detected samples, and  $Nfn$  is the number of false samples.

Table 2 shows the classification results of the proposed CNN models utilised for classification purposes in comparison to certain sophisticated methods.

In addition, as shown in Figure 5, we present the base template and the simulation environment of our technique

```

Input: Image label pairs  $\{(X_i, Y_i)\}_{i=1-N}$ , Bag-label pairs  $\{(B_i, Y_{i-ins})\}_{i=1-N}$ 
Training:
For  $n=1$  to  $N$  do
    For  $s=1$  to  $S$  do
        Multi-scale Transform:  $G_s: \{(X_i, Y_i)\}_{i=1-N}$ 
For  $b=1$  to  $B_n$  do
    Feature extractor:  $G_{features}: \{(B_{i-s}, Y_{i-ins})\}_{i=1-N}$ 
     $y_{ij} = \arg \max \{P(y_{ij} = k)\}$ 
Training end till last epoch
End
    
```

ALGORITHM 1: The Proposed Algorithm.

TABLE 2: Comparison of classification results of the proposed CNN model with sophisticated methods

Classification methods	Accuracy	Recall	Precision
DenseNet121	85.71	83.23	83.23
ResNet50	54.67	87.23	65.54
MobileNet	76.45	74.23	67.45
VGG16	83.45	48.34	73.56
MVNN	68.45	58.45	84.45
EfficientNet	46.53	65.45	45.67
Proposed method	<b>92.34</b>	<b>89.45</b>	<b>90.67</b>

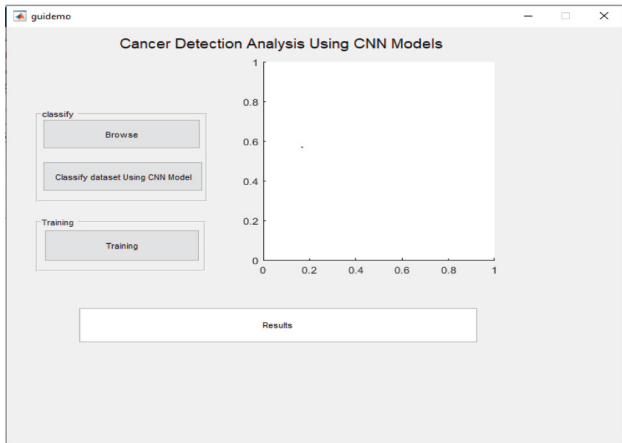


FIGURE 5: Cancer detection analysis template layout.

and the competing methods to visualise classification performance.

Figure 6 shows the training processing of the dataset when fed to the model. It will undergo the training of the model and classification according to the process by convolutional neural network layers. The results will be shown in the result tab as well which is showing a kind of cancer detected and the type of cancer is keratosis. In the same manner, when the complete dataset is processed for the training of the model, it may show the positive and negative parameters.

In Figure 7, it has been shown that the input images come out to be detected as 100 percent cancer cell means it is true given the output as 1.

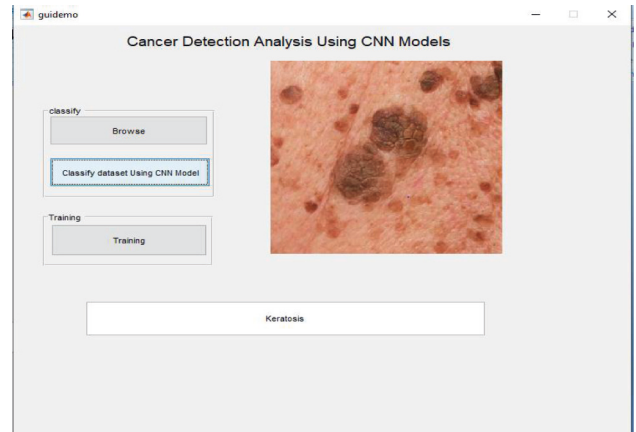


FIGURE 6: Cancer detected and classified after the testing phase.

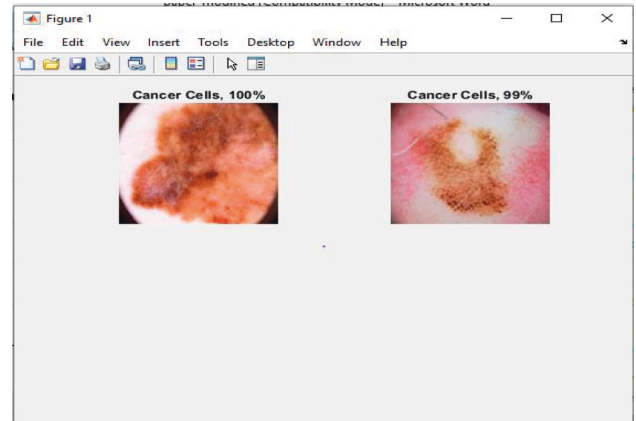


FIGURE 7: Cancer cells detected and shown the results with the % age of cancer cells.

Figure 8 shows the epoch processing time, as we are using the CPU of our laptop in the present work, which takes much time for the processing, so we stop the processing in between. The obtained result accuracy can be improved if the same model is used in a GPU environment.

Figure 9 shows the output that comes out from the processing of different layers of the convolutional neural

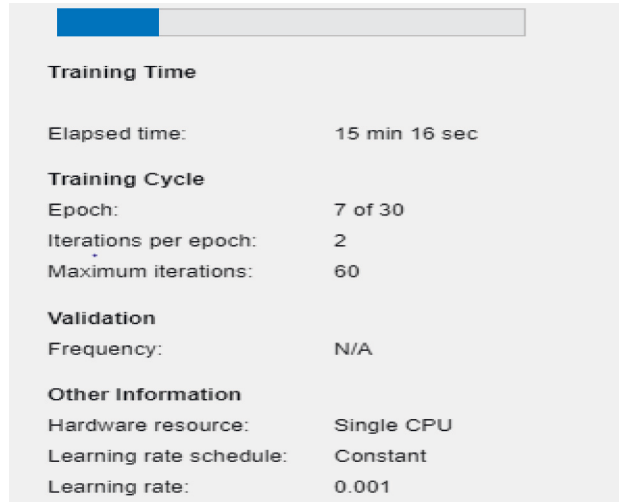


FIGURE 8: Intermittent processing of epoch/iteration.

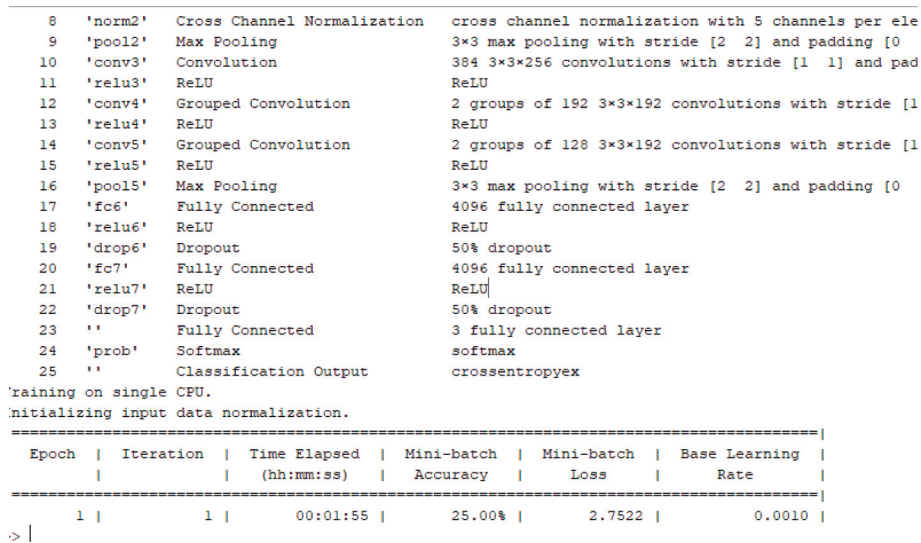


FIGURE 9: Convolutional output from different layers.

network. From iteration 1–7, the model takes for the initialization, and then from epoch 8 starts giving the output.

Figure 10 shows the processing analysis in terms of accuracy (%) and loss function. It clearly indicates that with the no. of iteration, the accuracy increases and the loss function decreases, which are an indication of the improved model.

Figures 11 and 12 show the different types of cancer cells detected and classified.

As already shown, Table 2 represents the calculated results from a different model and clearly indicates that the accuracy of the proposed model is much higher than the previous approaches along with better recall and precision values.

Finally, Figure 13 shows the graphical analysis of the precision with the recall values.

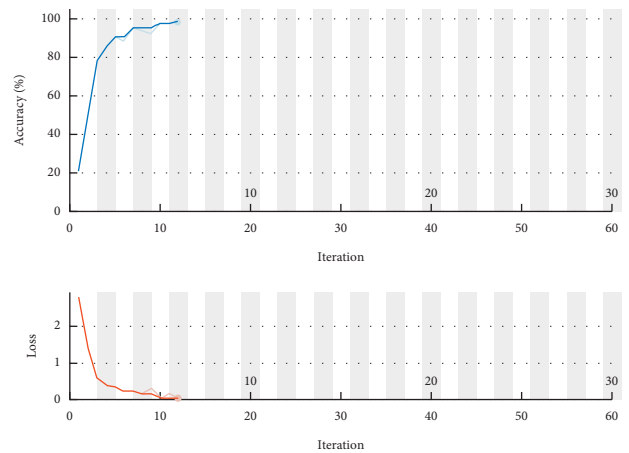


FIGURE 10: Performance analysis of accuracy (%) and loss function (%).

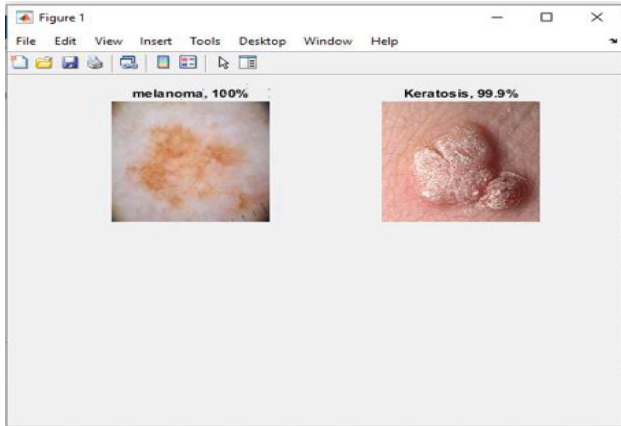


FIGURE 11: Cancer cells detected and classified.

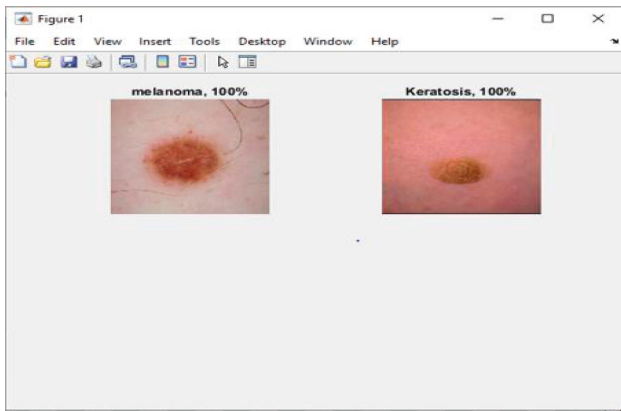


FIGURE 12: Cancer cells detected and classified.

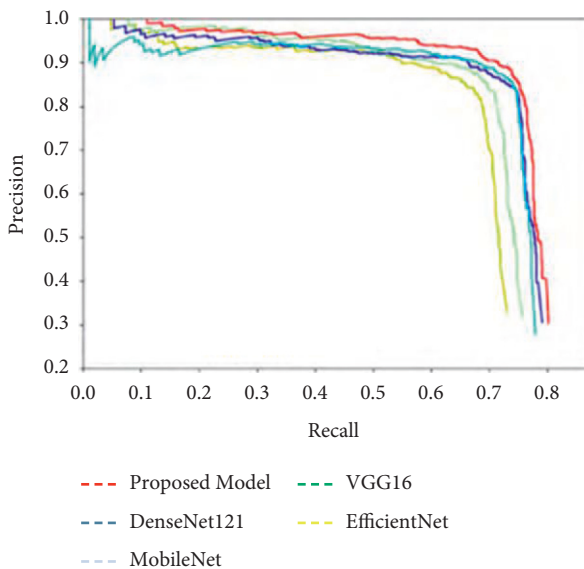


FIGURE 13: Analysis of precision with the recall values.

## 5. Conclusion and Future Scope

A multiconvolutional neural network learning approach is used in the proposed work to categorise entire cancer cell philosophy into normal and dangerous ones. Unlike the low-resolution images in the public dataset, the dataset used in our research work is a collection of samples collected in recent years using sophisticated technologies. The picture quality is assured, and it is more in accordance with contemporary clinical practice. For multiview mammography classification, we suggested a neural network based on global and local characteristics. Our model accepts both and shows better results in terms of accuracy, precision, and recall values. Given the limitations of the current study, we have indicated a few areas for future research. First, considering the importance of dataset size, we can investigate a larger dataset. Second, it is possible to explore integrating the annotations to improve accuracy. Third, multimodal analysis is an excellent way for any cancer screening and joint diagnosis integrating ultrasound and MRI which can be further analysed in future work.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The author declares no conflicts of interest.

### References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [2] M. Jansen and N. A. Wright, *Stem Cells, Pre-neoplasia, and Early Cancer of the Upper Gastrointestinal Tract*, Springer, Berlin, Germany, 2016.
- [3] M. Shaban, R. Awan, M. M. Fraz, A. Azam, and W. Tsang, "Context-aware convolutional neural network for grading of colorectal cancer histology images," *IEEE Transactions on Medical Imaging*, 2020.
- [4] A. Belsare, M. Mushrif, and M. Pangarkar, "Breast epithelial duct region segmentation using intuitionistic fuzzy based multi-texture image map," in *Proceedings of the 14th IEEE India Council International Conference (INDICON)*, IEEE, Roorkee, India, December 2017.
- [5] Y. Liu, F. Zhang, Q. Zhang, S. Wang, and Y. Wang, "Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 2020.
- [6] D. Gavrilov, L. Lazarenko, and E. Zakirov, "AI recognition in skin pathologies detection," in *Proceedings of the International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI)*, September 2019.

- [7] I. Sirazitdinov, K. Maksym, K. Ramil, and I. Bulat, "Data augmentation for chest pathologies classification," in *Proceedings of the IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, April 2019.
- [8] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, and E. Konen, "Chest pathology detection using deep learning with non-medical training," in *Proceedings of the IEEE 12th international symposium on biomedical imaging (ISBI)*, April 2015.
- [9] Y. Anavi, I. Kogan, E. Gelbart, and O. Geva, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," in *Proceedings of the 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, August 2015.
- [10] H. Nasir Khan, A. R. Shahid, B. Raza, A. H. Dar, and H. Alquhayz, "Multi-view feature fusion based four views model for mammogram classification using convolutional neural network," *IEEE Access*, vol. 7, Article ID 165724, 2019.
- [11] L. Sun, J. Wang, Z. Hu, Y. Xu, and Z. Cui, "Multi-view convolutional neural networks for mammographic image classification," *IEEE Access*, vol. 7, no. 99, p. 1, 2019.
- [12] K. J. Geras, S. Wolfson, S. G. Kim, L. Moy, and K. Cho, "High-resolution breast cancer screening with multi-view deep convolutional neural networks," 2017, <https://arxiv.org/abs/1703.07047>.
- [13] N. Wu, J. Phang, J. Park et al., "Deep neural networks improve radiologists' performance in breast cancer screening," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1184–1194, 2020.
- [14] C. Songsaeng, P. Woodtichartpreecha, and S. Chaichulee, "Multi-scale convolutional neural networks for classification of digital mammograms with breast calcifications," *IEEE Access*, vol. 9, Article ID 114741, 2021.
- [15] H. B. Ji, J. H. Park, H. P. Jin, and M. H. Sunwoo, "Multi-channel input deep convolutional neural network for mammogram diagnosis," in *Proceedings of the International SoC Design Conference (ISODC)*, Yeosu, Korea (South), October 2020.
- [16] C. Zhang, J. Zhao, J. Niu, and D. Li, "New convolutional neural network model for screening and diagnosis of mammograms," *PLoS One*, vol. 15, no. 8, Article ID e0237674, 2020.
- [17] H. D. Quy, V. K. Cao, H. Anh, and N. N. Son, "Multi-view digi 630 tal mammography mass classification: a convolutional neural network model approach," in *Proceedings of the International Symposium on Electrical and Electronics Engineering (ISEE)*, 2021.
- [18] A. Lx, Z. A. Lei, B. Hh, and Y. A. Zhang, "Neural networks model based on an automated multi-scale method for mammogram classification," *Knowledge-Based Systems*, vol. 208, 2020.
- [19] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft coco: common objects in context," in *Proceedings of the European Conference on Computer Vision*, pp. 740–755, Springer, Zurich, Switzerland, September, 2014.
- [20] D. Jia, D. Wei, R. Socher, L. J. Li, L. Kai, and F. F. Li, "Imagenet: a large-scale hierarchical image database," vol. 645, pp. 248–255, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 645, pp. 248–255, IEEE, Miami, FL, USA, June 2009.
- [21] S. Xie, R. Girshick, P. Doll'ar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July, 2016.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on 655 Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [23] D. Mahajan, R. Girshick, V. Ramanathan et al., "Exploring the limits of weakly supervised pretraining," in *Proceedings of the European Conference on Computer Vision*, pp. 181–196, Munich, Germany, September, 2018.
- [24] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and N. Houlsby, "An image is worth 16x16 words: transformers for image recognition at scale," 2010, <https://arxiv.org/abs/2010.11929>.
- [25] K. Han, Y. Wang, H. Chen et al., "A survey on visual transformer," 2020, <https://arxiv.org/abs/2012.12556>.
- [26] W. Wang, C. Chen, M. Ding, J. Li, and S. Zha, "Transbts: multi-modal brain tumor segmentation using transformer," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 109–119, Springer, 2021.
- [27] C. Zhang, J. C. Platt, and P. A. Viola, "Multiple instance boosting for object detection," in *Advances in Neural Information Processing Systems*, 2006.
- [28] Y.-L. Zhang and Z.-H. Zhou, "Multi-Instance learning with key instance shift," *IJCAI*, 2017.
- [29] S. Li, Y. Liu, X. Sui, C. Chen, and G. Tjio, "Multi-instance multi-scale CNN for medical image classification," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, October 2019.
- [30] M. Sun, T. X. Han, and M. C. Liu, "Multiple instance learning convolutional neural networks for object recognition," in *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR)*, December 2016.
- [31] C. Songsaeng, P. Woodtichartpreecha, and S. Chaichulee, "Multi-scale convolutional neural networks for classification of digital mammograms with breast calcifications," *IEEE Access*, vol. 9, Article ID 114741, 2021.
- [32] Z. Wang, C. Wang, J. Wei, and J. Liu, "Multi-class feature selection by exploring reliable class correlation," *Knowledge-Based Systems*, vol. 230, Article ID 107377, 2021.
- [33] Y. Liu, C. Wang, and Y. Wei, "Hippocampus segmentation in MR image based on atlas registration and broad learning," in *Proceedings of the Chinese Control And Decision Conference (CCDC)*, June, 2019.
- [34] T. M. Buzug, "Computed tomography," in *Springer Handbook of Medical Technology*, pp. 311–342, Springer, Berlin, Germany, 2011.
- [35] B. E. Warren, *X-ray Diffraction*, Courier Corporation, Massachusetts, MA, USA, 1990.
- [36] G. Campanella, M. G. Hanna, L. Geneslaw et al., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.
- [37] M. M. Dundar, B. Sunil, and C. R. Vikas, "A multiple instance learning approach toward optimal classification of pathology slides," in *Proceedings of the International Conference on Pattern Recognition*, August 2010.
- [38] Z.-H. Zhou and J.-M. Xu, "On the relation between multi-instance learning and semi-supervised learning," in *Proceedings of the 24th international conference on Machine learning*, 2007.
- [39] Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li, "Multi-instance learning by treating instances as non-iid samples," in *Proceedings of the*



- 26th annual international conference on machine learning, 2009.
- [40] S. Yin, Q. Peng, H. Li et al., "Multi-instance deep learning with graph convolutional neural networks for diagnosis of kidney diseases using ultrasound imaging," in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures*, pp. 146–154, Springer, Berlin, Germany, 2019.
- [41] C. P. Chen and Z. Liu, "Broad learning system: an effective and efficient incremental learning system without the need for deep architecture," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 10–24, 2017.
- [42] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.
- [43] M. Kandemir, A. Feuchitnger, and A. Walch, "Digital pathology: multiple instance learning can detect Barrett's cancer," in *Proceedings of the IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, April 2014.
- [44] A. T. Pham, R. Raich, and X. Z. Fern, "Dynamic programming for instance annotation in multi-instance multi-label learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2381–2394, 2017.
- [45] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163–180, 1994.
- [46] J. Fan, X. Wang, X. Wang, J. Zhao, and X. Liu, "Incremental wishart broad learning system for fast PolSAR image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1854–1858, 2019.
- [47] J.-W. Jin and C. L. Philip Chen, "Regularized robust Broad Learning System for uncertain data modeling," *Neurocomputing*, vol. 322, pp. 58–69, 2018.
- [48] M. Ye and J. Shen, "Probabilistic structural latent representation for unsupervised embedding," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020.
- [49] Z. Lu and I. C. Oliveira, "An efficient coding theorem via probabilistic representations and its applications," in *Proceedings of the 48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, 2021.
- [50] R. Lan, Y. Zhou, and Z. Liu, "Prior knowledge-based probabilistic collaborative representation for visual recognition," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1498–1508, 2018.
- [51] S. Cai, L. Zhang, and W. Zuo, "A probabilistic collaborative representation based approach for pattern classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, June 2016.
- [52] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," *Advances in Neural Information Processing*, vol. 15, 2002.
- [53] X.-S. Wei, J. Wu, and Z.-H. Zhou, "Scalable algorithms for multi-instance learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 4, pp. 975–987, 2016.
- [54] C.-C. Chang and C.-J. Lin, "Libsvm," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [55] L. Gan, J. Xia, P. Du, and J. Chanussot, "Multiple feature kernel sparse representation classifier for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5343–5356, 2018.
- [56] M. Sandler, A. Howard, and M. Zhu, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, June, 2018.
- [57] C. Zhang, J. Platt, and P. Viola, "Multiple instance boosting for object detection," *Advances in Neural Information Processing Systems*, vol. 18, pp. 1417–1424, 2005.
- [58] A. Bhandari, B. K. Tripathy, K. Jawad, and S. Bhatia, "Mohammad khalid imam rahmani, arwa mashat, cancer detection and prediction using genetic algorithms," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1871841, 18 pages, 2022.
- [59] S. Safdar, M. Rizwan, T. R. Gadekallu et al., "Bio-Imaging-based machine learning algorithm for breast cancer detection," *Diagnostics*, vol. 12, no. 5, 2022.