4th International Conference on Electrical Information and Communication Technology (EICT), 20-22 December 2019, Khulna, Bangladesh

Invasive Ductal Carcinoma Detection by A Gated Recurrent Unit Network with Self Attention

Ananna Biswas, Zabir Al Nazi, Tasnim Azad Abir
Dept. of Electronics and Communication Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh.
ananna9265@gmail.com, zabiralnazi@yahoo.com, tasnim.abir@ece.kuet.ac.bd

Abstract—**Representing around 80% of breast cancer, Invasive Ductal Carcinoma is the most common type of breast cancer. In this work, we have proposed a self-attention GRU model to detect Invasive Ductal Carcinoma. Self-attention is a way to motivate the architecture paying the attention to different locations of the sequence generated by an image effectively mapping regions of the image. The model was used to discriminate between cancerous samples and non-cancerous samples through training on the breast cancer specimens. The ability of discriminative representation has been improved using the self-attention mechanism. We have achieved the best average accuracy of 86%, a mean f1 score of 86% from our proposed model (*It should be noted that we used 1:1 train-test split to achieve this score*). We also experimented with a baseline CNN, ResNets (ResNet-18, ResNet-34, ResNet-50) and RNN variants (LSTM, LSTM + Attention). Our simple recurrent architectures with the attention mechanism outperformed Convolutional Networks which are traditional choices for image classification tasks. We have demonstrated how the scale of data can play a big role in model selection by studying different RNN, CNN variations for breast cancer detection scheme. This result is expected to be helpful in the early detection of breast cancer.**

Index Terms—**Gated Recurrent Unit, Recurrent Neural Network, Self Attention, Invasive Ductal Carcinoma**

I. INTRODUCTION

In recent years, deep learning has shown promising performance in different domains ranging from biomedical to medical diagnosis [1]. Deep convolutional neural networks have wide applications in medical disease classification. DCNN performs significantly good when trained with an optimized hyperparameter set but it needs a significant amount of data to train. Though RNN variants have shown reliable performances in image classification tasks, there are limited research works in this sector [2]. So, we have proposed an RNN-variant model for the detection of Invasive Ductal Carcinoma.

Breast cancer is the most common cancer in women worldwide. Invasive Ductal Carcinoma represents around 80 percent of the breast cancer types. It is one of the top reasons for woman cancer mortality. Early detection and accurate identification of cancer type can facilitate early diagnosis and timely treatment of breast cancer which can reduce the rate of deaths. Deaths of half a million breast cancer patients have already taken place and nearly 1.7 million new cases are arising per year. These numbers are expected to increase significantly in the upcoming years. So, automatic detection of breast cancer is an important step toward diagnosis.

There are related approaches in this domain to detect breast cancer with deep learning. The authors used a semi-automated segmentation method to characterize all microcalcifications in [3]. A discrimination classifier model was constructed to classify breast lesions based on microcalcifications and breast masses. They compared the performances of SVM, KNN, LDA and DL models and Deep learning-based approaches showed superior results. In [4], researches proposed an end-to-end recognition method by a novel CSDCNN model. Augmentation was applied to the BreaKHis dataset to boost the performance of the classifier. The authors proposed a convolutional neural network-based scheme for the classification of hematoxylin and eosin(HE) stained breast specimens in [5]. Multiple feature lists were explored with a sliding window approach for WSI classification which achieved an area under ROC of 0.92. In [6], authors proposed an Inception Recurrent Residual Convolutional Neural Network (IRRCNN) for breast cancer classification. BreakHis and Breast Cancer Classification Challenge 2015 datasets were used for image-based and patch-based evaluation. A CNN model with high generalized accuracy and minimal complexity was used to detect Invasive Ductal Carcinoma (IDC), Malignant and Benign tumors from histopathology and textual image datasets in [7]. SVM, Decision Tree, Logistic Regression and KNN were used to compare the accuracy among them. In [8], researchers proposed a CNN model based on the extraction of image patches to classify breast cancer histopathological images from BreakHis. In the paper, model adaptation was avoided for simplifying the model architecture and reducing computational costs. Some investigations had been done for achieving a high recognition rate. An effective Deep CNN method was used for the classification of HE stained histological breast cancer images. Data augmentation and feature extraction had been done for increasing the robustness of the classifier.

In this work, we have proposed a Gated Recurrent Unit with additive attention for the detection of Invasive Ductal Carcinoma. The main goal of our work is to claim that a Gated Recurrent Unit (special RNN) can also be used effectively in the particular fields of the image classification rather than CNN. For the reliability of our claim, we have compared

978-1-7281-6040-5/19/$31.00 ©2019 IEEE

different deep neural networks from CNN to RNN for specific model structures which are described in the subsection named Deep learning approaches with Convolutional and Recurrent Networks of the methodology section. We have also compared our model with other approaches which are shown in Fig 8. In the result analysis section, we have discussed our results and compared findings for the justification of our claim. Furthermore, the conclusion has comprised of summary and future objectives of our work.

II. METHODOLOGY

Deep learning has shown tremendous success in various domains like image processing, computer vision, medical imaging, natural language processing and many others [9], [10]. But there are limited resources of RNN based image classification. So, we have decided to adopt a relatively new approach- Gated Recurrent Unit with Attention mechanism to detect the invasive ductal carcinoma.

A. Dataset

In this article, Breast Histopathology Images were used for the detection of Invasive Ductal Carcinoma (IDC). The datasets were collected from [11]. 162 whole mount slide breast cancer specimens formed the original dataset that scanned at 40x. There were 277,524 patches of size 50 × 50 where 198,738 images were IDC negative and 78,786 images were IDC positive.

B. Pre-processing

The images are normalized prior to training. A sample of the images from each class has been shown in fig 1.

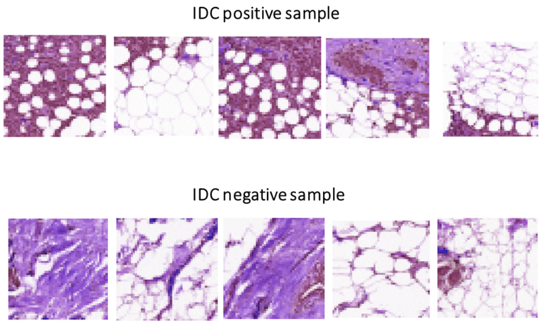

Fig. 1. The samples of histology patches of IDC positive and negative images

The dataset was split into training and testing subsets, with a ratio of 50:50 while 15% of training data were used for the validation. After normalizing, images are randomly shuffled. After shuffling, the shape of the training input tensor has been converted from (137162, 50, 50, 3) to (137162, 2500, 3) and the testing input tensor from (137161, 50, 50, 3) to (137161, 2500, 3). This is how the image can be represented as signal which is shown in fig 2.

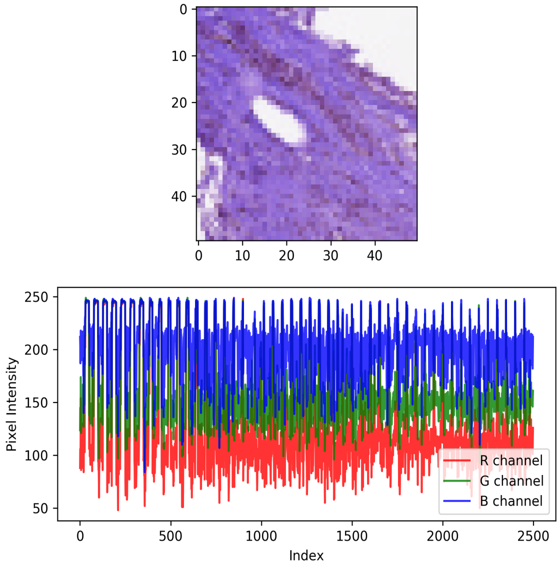

Fig. 2. IDC positive image and transformed signal representation

C. Deep learning approaches with Convolutional and Recurrent Networks

Generally, CNN is the first choice for image classification including medical image analysis as CNN has hierarchical feature extraction ability which reduces the need for synthesizing a separate feature extractor [12]. It helps the convolutional layers to learn effectively for classifying the images. In this regard, Residual Neural Network (ResNet) is an updated version of CNN that can train a deep neural network with 150+ layers using the skip connection strategy. On the contrary, RNN has quite a similar ability to recognize image features handling sequential data across time [13]. A few research works can be found in the RNN based image classification though it has shown superior performances in many sequential tasks primarily. So, in this paper, we have considered using the RNN variants with an attention mechanism for classifying images that have shown promising performance.

RNN suffers from gradient exploding and vanishing problems while LSTM can solve the problems using various memory cells replacing the hidden units that can control the flow of the information and reduces long-term dependencies of the input data. But LSTM has some limitations of having a large number of parameters compared to RNN which increases computational complexity. In this case, Gated Recurrent Unit (GRU) is a good option which has the ability to reduce the number of gates in LSTM. Moreover, GRU controls the flow of information in the same way as LSTM without using a memory unit which makes GRU less complex compared to LSTM.

So, we have used GRU over LSTM for better computational efficiency and faster training ability. Here, we have also used

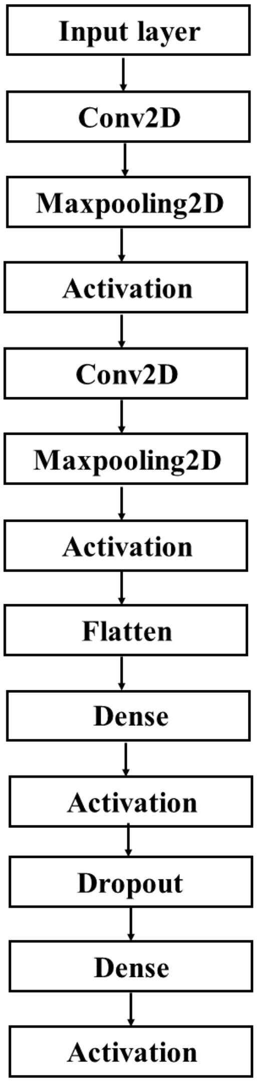

Fig. 3. Baseline CNN Architecture

additive attention with GRU that extends Neural Network's capabilities in various predictions. Using self attention, RNNs focus on a specific part of a subset of the given input data. At every time steps, it focuses on different positions of the data in the inputs. The attention mechanism improves the GRU network's ability for discriminative representation. The purpose of the attention learning mechanism is to exploit the intrinsic self-attention ability of GRU that has enhanced the performances of the model in image classification. To classify images, we have evaluated different neural architectures. They are CNN, ResNet18, ResNet34, ResNet50, LSTM, LSTM + attention, and GRU + attention. The architecture of each network is described below.

- **Baseline Convolutional Neural Network:** The baseline Convolutional Neural Network (CNN) consists of two convolutional layers, two pooling layers, and two fully connected layers shown in fig 3. ReLU has been used as activation in intermediate layers, and sigmoid in final dense layer.
  Input layer loads input data and feeds to convolutional layers. Our input image size was 50-by-50 where the number of channels was 3. We have used two convolutional layers that produce a set of filters from the input data. There is one pooling layer after each convolutional layer. The pooling layers downsample the spatial dimension of the input. Two types of activation functions have been used here that introduce non-linear properties to our network. Two dense layers feed all outputs from the previous layer. First dense layer has 50 neurons and second dense has 2 neurons in the network. The flatten layer took place between the fully connected layer and the convolutional layer. It transforms a two-dimensional matrix into a vector. The vector can be fed into a fully connected neural network classifier. For the sake of regularization and solving the overfitting problem in CNN dropout has been used.

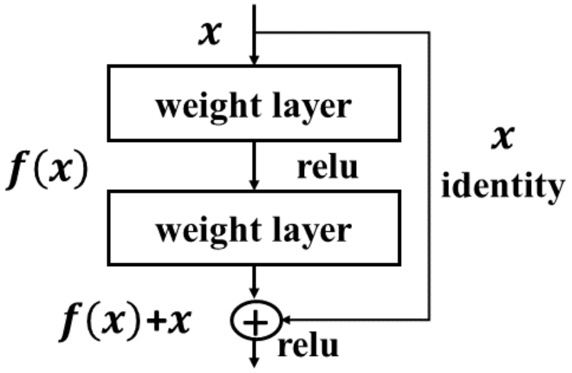

Fig. 4. ResNet Block

- **Residual Networks:** Resnet-18 is a convolutional network that exploits residual learning.CNN suffers from overfitting and optimization problems that increase training error. Residual Networks can train such deep networks through residual modules [14]. The Residual network follows the skip connection technique that simplifies the network. So, it learns better than a baseline CNN. We have experimented ResNet-18 with ResNet-34 and ResNet-50 where ResNet-18 has shown better performance. ResNet-18 has a basic block with the input and output layer. The basic block of a Residual Network is shown in fig 4. [15]
- **Long Short-Term Memory Network:** The LSTM is a Recurrent Neural Network (RNN) that can solve the problem of short-term memory. The LSTM network has three gates (forget gate, input gate, and output gate) and one cell state that regulates the flow of the information. There is a sigmoid function that is used to squishes values

between 0 and 1. The forget gate decides whether the information is important or not.

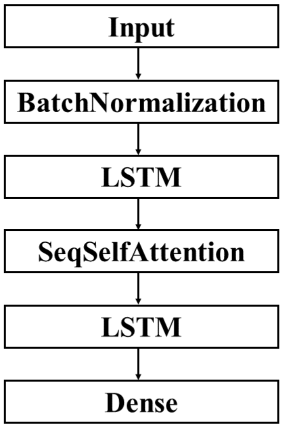

Fig. 5. LSTM+attention Architecture

Information from the current input and the previous hidden is passed to the next state through the sigmoid function. The output that is closer to 0 will be forgotten and the output that is closer to 1 will be stored. The duty of the input gate is to update all the state where the cell state works as a transport highway or as a memory. The output gate selects the next hidden state that is used for the prediction. The LSTM network which is used here for breast cancer detection is shown in fig 5. An attention mechanism is also added with this network for more accurate results in the detection of invasive ductal carcinoma.

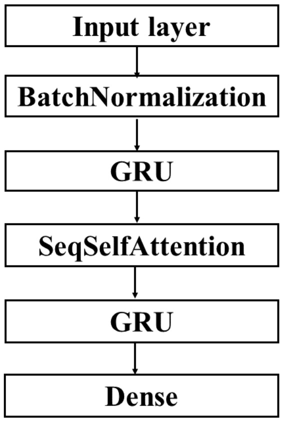

Fig. 6. GRU+attention Architecture

- **Gated Recurrent Unit Network:** The GRU is a new generation of Recurrent Neural Network (RNN) which has almost same functionality as that of the LSTM network. The GRU network has basically two gates (reset gate and update gate) and no cell sate. The functions of the different gates of the GRU are described below.

1. Update Gate: It determines the quantity of the past information needed to be passed along into the future. It is similar to the output gate of the LSTM network.
2. Reset Gate: It determines how much of past information have to forget. It works as the combination of the input gate and forget gate of the LSTM network.
3. Current Memory Gate: It is not regarded as an individual part rather a sub-part of the Reset Gate as it is incorporated into the reset gate. It helps to reduce the effect that past information has on the current information.
Though the functions of the two (GRU and LSTM) networks are quite similar, the network of the GRU network is less complex than the LSTM network [16]. In this research work, the GRU network with an attention mechanism has been used for the early detection of invasive ductal carcinoma which is shown in fig 6. The attention has focused on the discriminatory regions between the non-cancerous and cancerous images that helped to achieve better performance. In the processing of breast cancer data, the self-attention technique follows the context for each timestep.
Attention: We have used additive local attention [17]–[19].

$$\begin{aligned} h_{t,t'} &= tanh(x_t^T W_t \\ &\quad + x_t'^T W_x + b_t) \end{aligned} \tag{1}$$

$$e_{t,t'} = \sigma(W_a h_{t,t'} + b_a) \tag{2}$$

$$a_t = softmax(e_t) \tag{3}$$

$$l_t = \sum_{t'} a_{t,t'} x_{t'} \tag{4}$$

A GRU or LSTM layer was used as an encoder for the hidden state representations (h_t). The attention matrix A determines the similitude of any token with adjacent tokens from the input signal representations of the images. The similarity between the hidden state representations h_t and h_t' of tokens x_t and x_t' at timesteps t and t' are captured by attention element $a_{t,t'}$. The attention scheme is implemented based on equations [1-4], where W_t, and W_x denote the weight matrices for h_t and h_t' and W_a is for the representation of non-linear relationship for the hidden states; b_t and b_a are the bias vectors. The point-wise sigmoid operation has been a representation by σ. Finally, the attention hidden state representation l is calculated, where l_t is a token at timestep t which is calculated based on the weighted summation of h_t' of all other tokens at timestep t' and $a_{t,t'}$. l_t denotes the amount to attend to a token-based on their adjacent context and token importance.

TABLE I
PERFORMANCE COMPARISON

Model	Accuracy	Precision	Recall	F1 Score	Hyperparameters			
					Learning rate	Optimizer	Loss function	Activation function
CNN	0.73	0.52	0.73	0.61	0.05	Adam	categorical crossentropy	relu, softmax
ResNet-50	0.73	0.68	0.73	0.70	0.001	Adam	categorical crossentropy	relu, softmax
ResNet-34	0.77	0.75	0.77	0.76	0.001	Adam	categorical crossentropy	relu, softmax
ResNet-18	0.79	0.81	0.79	0.80	0.001	Adam	categorical crossentropy	relu, softmax
LSTM	0.82	0.81	0.82	0.81	0.001	RMSprop	categorical crossentropy	sigmoid
LSTM + attention	0.85	0.84	0.85	0.84	0.001	Adam	categorical crossentropy	sigmoid, softmax
GRU + attention	0.86	0.87	0.86	0.86	0.001	Adam	categorical crossentropy	sigmoid, softmax

III. Result Analysis

In this work, we compared CNN, ResNet and multiple RNN variants with our model. The proposed attention + GRU model outperformed other models while the numbers of hyperparameters were almost the same which is shown in Table 1. As the dataset contained good enough samples, we decided to train on only 50% of the data. We have used 50% of the input data (137162 samples) for the training and the rest 50% (137161 samples) for the testing. Most of the experiments were performed in Google Colab[1] where a maximum of 12 GB ram is allowed (12GB NVIDIA Tesla K80 GPU), so we had to use smaller set for training, but even after training the models on only 50% of the data, the performance was comparable to benchmark results. We have got a better estimate of our model accuracy with this test data. This 50:50 train/test split has given us a measure of the classifier's strength compared to other classifiers. From table 1, we can compare the performance of all the models used in the work. The baseline-CNN model and the Resnet-50 model have shown quite similar performances where the ResNet-18 model has shown better performance than the ResNet-34 and the ResNet-50 model.

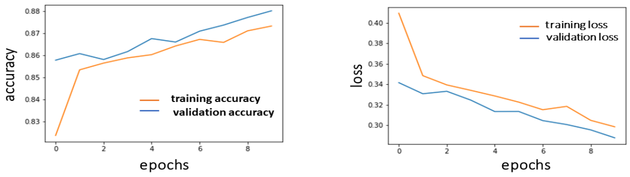

Fig. 7. Training curve of the attention RNN-GRU model

For the ResNet-50 model, the normalized values have to pass many of the layers where our image size was 50 by 50 pixels. So, due to low spatial resolution and smaller training split deeper networks were not able to learn most of the useful features resulting in poor accuracy (0.73) and f1 score (0.70). We can observe this pattern in table 1, as the CNN model gets deeper the performance drops. For solving this problem, we have used the LSTM model that has achieved better accuracy (0.82) and f1 score (0.81) than the ResNet-18 model. Furthermore, we have added a self-attention layer for exploiting the intrinsic self-attention ability of LSTM that increases both accuracy (0.85) and f1 score (0.84). But LSTM has more computational complexity which takes a longer time than other RNN variants. So, we have explored GRU with an attention layer for making the model simple that decreases the training time. We get the training curve of the model in fig 7.

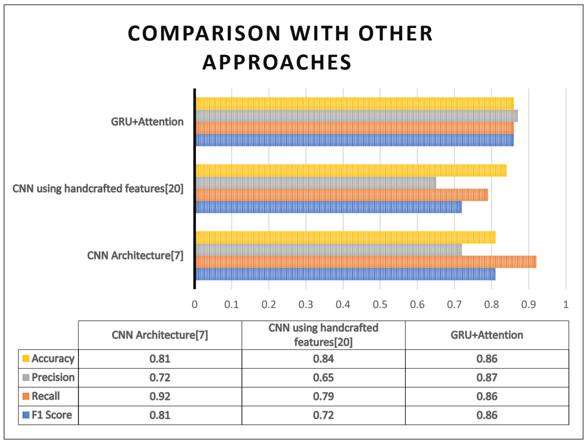

Fig. 8. Comparative performance analysis

We have used 15% of the training data for the validation. Adam optimizer, categorical_cross-entropy loss function, and sigmoid and softmax activation (Attention) functions have been used here. During training our data we have used regularization which is not used in our validation part. Generally, the training accuracy is greater than the validation accuracy which is quite common to all. But in this case, validation accuracy is greater than training accuracy due to the regularization. When we used dropout in training- disabling some neurons, some of the information about each sample has lost. So, we have discovered the low performance of the training than validation- where dropout has not been used. Overall, the GRU + attention model has achieved a promising performance (accuracy=0.86 and f1 score=0.86). As the spatial dimension is limited in the data, the results suggest the images can be represented as 1-dimensional signals and still be classified with good performance.

Fig.8 illustrates the performance comparison of different models. In [7], CNN architecture has been used in the breast cancer classification using histopathological images and achieved 81% accuracy. Handcrafted features have been used in [20] with the CNN model to detect invasive ductal carcinoma where 84% accuracy has been achieved. In this case, our proposed architecture has achieved 86% accuracy which is promising.

[1] https://colab.research.google.com/

IV. Conclusion

In this paper, we have proposed an attention + GRU network for classifying images that perform better than CNN and RNN variants. We have experimented with different CNN and RNN models with similar hyperparameters that are chosen very carefully for the training. The main contribution of our research work is that we have tried to investigate the effect of spatial dimension on model selection and showed that an efficiently trainable Gated Recurrent Unit with Attention can outperform traditional neural networks. We have used the attention mechanism with our proposed model that can focus on the specific part of the input data. The attention enhanced the performance of the model and improved learning for the transformed data. The model has less computational complexity than other RNN variants that reduces the training time which is an important aspect of the real-time image diagnosis. We are yet to explore attention for explaining the predictions and the optimization of the hyperparameters for improving the performances in the IDC classification task.

References

[1] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.

[2] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.

[3] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, "Discrimination of breast cancer with microcalcifications on mammography by deep learning," *Scientific reports*, vol. 6, p. 27327, 2016.

[4] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," *Scientific reports*, vol. 7, no. 1, p. 4172, 2017.

[5] B. E. Bejnordi, J. Lin, B. Glass, M. Mullooly, G. L. Gierach, M. E. Sherman, N. Karssemeijer, J. Van Der Laak, and A. H. Beck, "Deep learning-based assessment of tumor-associated stroma for diagnosing breast cancer in histopathology images," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 929–932.

[6] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *Journal of digital imaging*, pp. 1–13, 2019.

[7] P. Mohapatra, B. Panda, and S. Swain, "Enhancing histopathological breast cancer image classification using deep learning," vol. 8, 06 2019.

[8] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *2016 international joint conference on neural networks (IJCNN)*. IEEE, 2016, pp. 2560–2567.

[9] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, 2019.

[10] Z. Al Nazi and T. A. Abir, "Automatic skin lesion segmentation and melanoma detection: Transfer learning approach with u-net and dcnn-svm," in *Proceedings of International Joint Conference on Computational Intelligence*. Springer, 2019, pp. 371–381.

[11] P. Mooney, "Breast histopathology images," *http://spie.org/Publications/Proceedings/Paper/10.1117/12.2043872*, 2017.

[12] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *arXiv preprint arXiv:1901.06032*, 2019.

[13] B. Chandra and R. K. Sharma, "On improving recurrent neural network for image classification," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 1904–1907.

[14] M. R. Mamun, Z. Al Nazi, and M. S. U. Yusuf, "Bangla handwritten digit recognition approach with an ensemble of deep residual networks," in *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*. IEEE, 2018, pp. 1–4.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *International Conference on Machine Learning*, 2015, pp. 2342–2350.

[17] G. Zheng, S. Mukherjee, X. L. Dong, and F. Li, "Opentag: Open attribute value extraction from product profiles," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 1049–1058.

[18] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[19] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[20] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, and A. Madabhushi, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *Medical Imaging 2014: Digital Pathology*, vol. 9041. International Society for Optics and Photonics, 2014, p. 904103.