Automatic Chromosome Classification using Deep Attention Based Sequence Learning of Chromosome Bands

Monika Sharma[*], Swati[*], Lovekesh Vig
TCS Research, New Delhi, India
Email : { monika.sharma1 | j.swati | lovekesh.vig }@tcs.com

Abstract

*Chromosome Karyotyping refers to the task of segmenting and classifying individual chromosome images obtained from stained cell images microphotographed during the metaphase stage of cell division. The karyotyped images are useful for diagnosis of genetic disorders such as down syndrome, turner syndrome and certain types of cancers. In many hospitals and labs, a significant amount of manual effort and time is spent on segmenting and classifying the individual chromosome images. Recently, deep learning models have been applied to automate this task with promising results. An important characteristic of a chromosome is the presence of sequence of dark and light bands produced by giemsa staining which is used by cytogeneticists to manually perform karyotyping. We propose Residual Convolutional Recurrent Attention Neural Network (Res-CRANN) which exploits this property of band sequence for chromosome classification. Res-CRANN is end-to-end trainable in which a sequence of feature vectors, extracted from the feature maps produced by convolutional layers of Residual neural networks (ResNet) is fed into Recurrent Neural Networks (RNN) and subsequently, an attention mechanism is applied on top of RNN output sequences which are further classified into one of the 24 labels. The attention mechanism after recurrent layers facilitates the network to learn to pay selective attention to the sequence of bands and relate them to different classes of chromosomes. We demonstrate the proposed architecture's efficacy on a publicly available Bioimage chromosome classification dataset and observe that our model outperforms the baseline models created using traditional deep convolutional neural network and ResNet-50 by approximately **3**% Top-1 classification accuracy.*

1. Introduction

Cytogeneticists perform karyotyping which involves segmentation and classification of individual human chromosomes during metaphase stage of cell division. An example of metaphase image is shown in Figure 1. A healthy human cell has 23 pairs of chromosomes, out of which 22 are called autosomes and 23^{rd} pair is called sex chromosomes (X and Y). The chromosome images are giemsa-stained during metaphase stage of cell division to produce visible light and dark colored bands known as banding pattern (see Figure 2(a)). Doctors analyze these stained cell images on the basis of various properties of chromosomes like length, centromere position, banding pattern, etc. and categorize them into one of the 24 chromosome classes. The classified chromosomes of a human being are then represented in a standard image format called karyotype, as shown in Figure 2(b).

Figure 1. Cell spread image during metaphase stage.

Karyotyping is performed by doctors for the diagnosis of various birth defects, cancers and genetic disorders. Even after years of expertise, doctors spend considerable amount of time and manual effort to perform segmentation and classification. This motivates us to automate / semi-automate the process of karyotyping which would assist doctors and reduce their cognitive load while also saving their precious time spent during the process. Conventional Deep Con-

[*]indicates equal contribution arranged alphabetically

978-1-5090-6014-6/18/$31.00 ©2018 IEEE

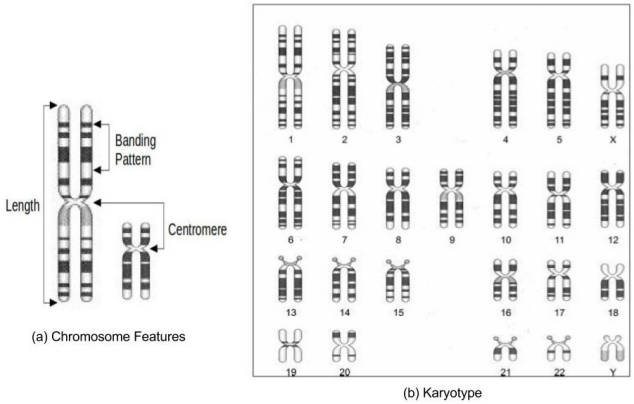

Figure 2. (a) Chromosome distinguishing features - length, centromere and banding pattern, and (b) A karyotype image.

volutional Neural Networks (DCNN) [20, 21] have proved to be the state-of-art techniques for wide variety of tasks in computer vision such as image classification[20], object recognition[10] and image captioning[33] over the past years. Recently, DCNN has also been applied to chromosome karyotyping with encouraging results [29, 18]. In this paper, we explore the state-of-the-art deep learning model, ResNet [14] for chromosome classification. Here, we make an assumption that we have segmented and straightened individual chromosomes available with us and we take up the task of automating chromosome classification to accelerate karyotyping.

We know that chromosomes have characteristic and unique sequence of light and dark colored bands, as shown in Figure 2(a), which help to distinguish between different chromosome classes. The position and size of bands are highly chromosome-specific. These bands are a useful way of subdividing various regions of chromosomes and each band is assigned a specific number that doctors use for assigning label to a particular chromosome. However, DCNNs are not able to exploit this characteristic properly as they mostly concentrate on learning image features and thus, cannot be directly applied to learn sequence prediction. We propose an architecture to learn sequence of appearance of these bands explicitly by employing Recurrent Neural Networks (RNN) which have shown their worth in processing sequential data of varied length, such as text or speech [12, 7, 32]. More specifically, we use Long-Short Term Memory Networks (LSTM) [16], a special type of RNN used for modeling long-term dependencies in the data using a gating mechanism. We use the combination of DCNN and LSTM in a similar fashion as proposed in Convolutional Recurrent Neural Networks (CRNN) [30] which has been applied for scene text recognition. The process of learning sequence of light and dark bands in chromosomes is analogous to reading characters one-by-one in image-based text sequence recognition task. Thus, in our method, we extract sequence of feature vectors by horizontal or row-wise concatenation of feature maps produced by convolutional layers of ResNet and feed this sequence to LSTM. Further, we augment the LSTM with attention mechanism to help the network learn to pay attention to the relevant regions of the input image or sequence [3, 24, 13]. We demonstrate that ResNet in combination with LSTM and attention improves the classification accuracy of chromosomes on a publicly available dataset [1, 27]. As a preprocessing step, we also normalize the individual chromosomes length across each patient and then the normalized chromosomes are sent as input to the proposed architecture. While exploring different techniques, we make the following main contributions in this paper:

- We propose an end-to-end trainable Residual Convolutional Recurrent Attention Neural Network (Res-CRANN) for automatic chromosome classification utilising the sequence of banding patterns in the image. The architecture of the network is shown in Figure 3.
- We benchmark the proposed Res-CRANN model against Deep Convolutional Network (DCNN) [29] and ResNet-50 [14] for the task of chromosome classification on a publicly available Bioimage Chromosome Classification dataset [1, 27].
- We observe that the addition of LSTM and attention layers to the ResNet-50 model improves the chromosome classification accuracy by approximately **3%**.
- We also visualize the saliency maps produced by Res-CRANN and observe that learning sequence of chromosome bands and adding attention to the network remove noise from the saliency maps as compared to the ones produced by ResNet-50 network.

The rest of the paper is organized as follows: Section 2 gives an overview of related work in the field of chromosome karyotyping. Section 3 describes the proposed methodology for automatic chromosome classification which is followed by a brief description of ResNet, CRNN and attention based sequence learning in Section 3.1, 3.2, 3.3. In Section 4, the proposed architecture of Res-CRANN is explained. Subsequently, in Section 5, we present details about the dataset, the training setup utilized and a discussion on the obtained results. Finally, we conclude the paper in Section 6.

2. Related Work

Chromosome karyotyping is a two-stage process: first is the segmentation of cell spread image during metaphase stage of cell division and second is the classification of individual chromosomal segments. Several attempts have been

made in the past for automation of karyotyping [5], owing to the laborious and time-consuming process and various efforts have been made to find computational methods for chromosome segmentation [23] and classification of chromosomes [26, 6, 29, 18]. As chromosomes are bent in different orientations after giemsa staining process, a lot of work has also been carried out on straightening of bent chromosomes [17, 28] as it improves the classification accuracy. However, in this paper, we make an assumption that we have already been provided with segmented and straightened individual chromosomes and our task is to automate the stage of chromosome classification.

In past, automation of chromosome classification was usually carried out by training artificial neural networks [11] on a set of chromosome features like length of chromosome, centromere position and banding profile extracted using traditional computer vision techniques [22]. However, recently with the boom in deep learning techniques, researchers have also demonstrated the efficacy of Deep CNN [29] and Siamese Networks where labeled data is scarce [18]. Since, deep neural networks are difficult to train, a Residual neural network (ResNet) [14] was proposed to ease the training procedure and this network has shown state-of-the-art results in various vision tasks like object detection, object localization and image segmentation. Therefore, in this paper, we explore ResNets for automatic chromosome classification.

However, to further improve the classification accuracy, we exploit the sequence of light and dark colored bands that appear on chromosomes after giemsa-staining process and are characteristic feature of different chromosomes. For this, we use Recurrent Neural Networks (RNN) which have shown impressive results on sequential data such as text and speech [12, 7, 32]. The input of the Recurrent Networks not just take the current example they see, but also what they have perceived previously in time as the sequential data have some hidden information and recurrent nets use their memory to perform tasks that feedforward networks cannot do. We particularly use Long Short Term Memory networks (LSTM), a form of recurrent network, to utilize the information hidden in the sequence of chromosome bands. We borrowed the concept of Convolutional Recurrent Neural Networks (CRNN) [30] used for scene based text recognition where the characters are read one-by-one in an image. We concatenate the feature maps produced by convolutional layers of ResNet along rows to form a sequence of feature vectors which are then fed into LSTM to learn the sequence of chromosome bands of different chromosome classes. Additionally, the attention networks [3, 24, 13] have shown the state-of-the-art performance for various tasks like text-recognition, object detection and localization. Thus, we employ the attention mechanism after recurrent layers to help the network learn to concentrate on region of interest, i.e., sequence of bands.

3. Proposed Methodology

This section describes the proposed method for chromosome classification as shown in Figure 3. As we know cytogeneticists classify chromosomes into different types based on three main distinguishing properties, namely, length of chromosomes, centromere position, and sequence of dark and light colored bands. To preserve the length and centromere position of chromosomes across different patients, we normalize the length of chromosome segments across each patient before passing to the neural network. Deep convolutional neural networks (DCNN) mostly concentrate on learning image features and are unable to concentrate on learning sequence of chromosome bands. However, the size and position of the chromosome bands are specific to different type of chromosomes. Hence, to explicitly make the network learn the sequence of bands, we use Recurrent Neural Networks (RNN). More specifically, we use Long Short Term Networks (LSTM - a variant of RNN) on top of convolutional feature maps produced from underlying DCNN processed in a similar fashion to Convolutional Recurrent Neural Networks (CRNN)[30]. This learning of sequence of chromosome bands is analogous to reading of characters. Since, the length of sequences generated from CRNN is quite large, we employ attention mechanism on sequence learning to make the network learn to pay attention to the relevant part of the sequence. Therefore, our proposed network architecture consists of a combination of Residual Neural Networks (ResNets) used as DCNN block followed by LSTM and Attention block to learn the features based on sequence of chromosome bands. Thus, we named our network as Residual Convolutional Recurrent Attention Neural Network (Res-CRANN) whose architecture is explained in Section 4. Each essential block used in this architecture is described in the following subsections:

3.1. Residual Neural Networks

Residual neural networks, or ResNets [14] are very intriguing and proven to be very robust in various visual and non-visual tasks. It utilizes the deep residual learning framework which makes the training of deeper neural networks easier. This residual framework replaces the blocks of convolutional layers by shortcut connections to form shortcut blocks, also known as residual blocks. The residual block, as shown in Figure 4, can be expressed as :

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \{\mathbf{W_i}\}) + \mathbf{x} \tag{1}$$

In this equation, $\mathbf{x}$ and $\mathbf{y}$ are input and output of the residual block. The function $\mathbf{F}(\mathbf{x}, \{\mathbf{W_i}\})$ represents the residual mapping to be learned. These blocks are trying to model an optimal function which is closer to an identity mapping than to a zero mapping, and that it should be

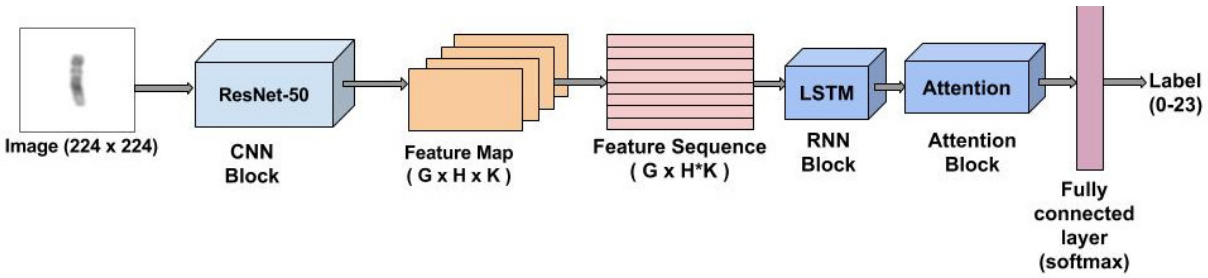

Figure 3. The proposed architecture of Res-CRANN for chromosome classification. The ResNet-50 model provides K feature maps of size $G \times H$ (G and H are height and width of a feature map) which are concatenated horizontally to produce feature sequence of dimension $G \times H * K$. This sequence is passed through LSTM followed by attention block and finally through fully connected softmax layer which outputs a class label in range (0 - 23) for the input chromosome image.

easier to find the perturbations with reference to an identity mapping than to a zero mapping. This simplifies the optimization of ResNets at almost no cost. Subsequent residual blocks in ResNets are thus responsible for fine-tuning the output of previous block, instead of generating the desired output from scratch.

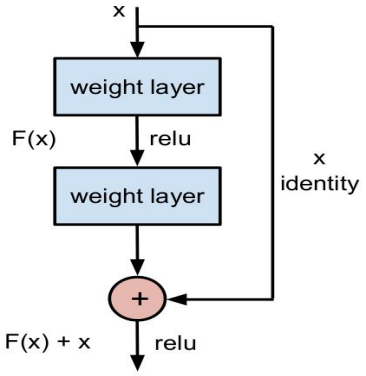

Figure 4. Residual block of ResNet - reproduced from [14]

The main property of these residual connections is that they neither introduce extra parameters nor increase computational complexity. The residual blocks are stacked and residual mapping helps in handling the vanishing gradient problem [31, 25] in the deeper networks which makes optimization easy and thus, results in improvement of training efficiency. ResNets have multiple versions of architectures, represented as m-layer ResNet model, where m is the number of layers in the corresponding architecture and can take values in $\{50, 101, 152\}$.

3.2. Convolutional Recurrent Neural Networks

Convolutional Recurrent Neural Networks (CRNN) [30] are a combination of Deep Convolutional Neural Networks (DCNN) and Recurrent Neural Network (RNN), whose architecture is mainly designed to recognize sequence-like objects in the image. The traditional architecture of CRNN involves convolutional layers, recurrent layers and transcription layers from bottom to top. In our implementation, we have replaced the transcription layers by softmax layer for the task of classification. The convolutional layers of underlying DCNN in CRNN architecture takes image as an input and produce corresponding feature maps. These feature maps are then used to extract sequence of feature vectors where each feature vector is generated by concatenating rows of filter maps from top to bottom . Subsequently, this sequence of feature vectors is fed to recurrent neural network, particularly long short term memory networks (LSTM) which produce labels for each frame of the feature sequence. A softmax layer at the top translates the per-frame predictions by LSTM into a class-label sequence. This entire network is trained end-to-end by using a single loss function which jointly trains both constituent networks (i.e., DCNN and LSTM).

3.3. Attention based Sequence Learning

The attention mechanism in the neural networks has become very popular in recent times. It is based on the visual attention mechanism found in humans. When a human sees an image, his visual attention boils down to focus on a certain region-of-interest in an image with *high resolution* while perceiving the surrounding area of the image with *low resolution*, and then he keeps adjusting his focus over time. This attention mechanism is being widely applied in various tasks ranging from machine translation[4], speech recognition[9], image captioning[34] and question answering[15]. In case of long sequences, attention is utilized to make the network learn to concentrate on particular region-of-interest of these sequences. The attention mechanism in neural networks reduces the number of parameters required to train the network and also avoids the chances of overfitting.

4. Architecture of Res-CRANN

The flow diagram of proposed architecture is shown in Figure 3. It is a deep attention based CRNN model

with bottom DCNN replaced by 50-layer ResNet model followed by LSTM whose output sequence is passed through attention module. We call this network architecture as Residual Convolutional Recurrent Attention Neural Networks (Res-CRANN). The complete network is trained end-to-end and the fully connected softmax layer at the output gives the class-label for a chromosome image.

The bottom of Res-CRANN is a 50-layer ResNet model which produces feature maps as a result of convolutional layers. These feature maps are then converted into a feature sequence in a similar fashion to CRNN[30] by concatenation of the horizontal vectors of each feature map whose mathematical interpretation is explained below :

$$F_g^i = concat(f_{gk}^i)\,; \quad k = 1\ to\ K \tag{2}$$

where, F_g is a feature vector of dimension $H * K$ obtained from ResNet convolutional feature maps, f_{gk} is g^{th} horizontal vector of k^{th} filter map, i represents the i^{th} image and K is the total number of filter maps. The final feature sequence S^i of the i^{th} image is given by :

$$S^i = [F_1^i, F_2^i, ..., F_G^i]^{G \times H * K} \tag{3}$$

where, G and H are the height and width of the feature map obtained from the topmost convolutional layer of ResNet-50 model.

The feature sequence S^i is then passed through LSTM which learns to predict the band sequence. Subsequently, the LSTM output is fed into an attention block which learns the attention weights for each of the feature vector in the sequence as described below:

Let M be the matrix given by $M = [m_1, m_2, .., m_N]$, where m_i represents the output vector sequence that the LSTM produces and N is the length of sequence. The output representation R of attention block is the weighted sum of these vectors, which is defined as :

$$L = tanh(M) \tag{4}$$

$$\alpha = softmax(w^T L) \tag{5}$$

$$R = M\alpha^T \tag{6}$$

where, w are the weights to be learned for attention block. The final representation r used for classification is given as follows :

$$r = tanh(R) \tag{7}$$

The representation r produced from the attention mechanism is then passed through the fully connected layer having *softmax* activation which classifies the image into its corresponding class-label.

5. Experiments

This section begins with the description of a publicly available Bioimage Chromosome Classification dataset in 5.1. Subsequently, 5.2 explains the experimental setup utilized to conduct our experiments. In Subsection 5.3, we discuss the performance of the experiments conducted and compare our results with the baseline models.

5.1. Dataset

We have evaluated performance of our proposed architecture on Bioimage Chromosome Classification dataset [1, 27] which is publicly available online. The dataset consists of total 5256 chromosome images of healthy patients which are manually segmented and classified by an expert cytogenecist in order to create the labeled dataset. Out of these 5256 images, we have utilized 4176 for training, 360 for validation and 720 for testing purpose. For conducting our experiments, the resolution of chromosome images is set to 224×224 in gray-scale. In addition, we have applied a pre-processing step of length normalization[29] to every chromosome image in the dataset.

5.2. Training Details

The first baseline method used for comparing the performance of our proposed method consists of traditional Deep CNN [29] (row 1 of Table 1). Deep CNN model was trained using Adam optimizer with learning rate of 10^{-4} and rest of the parameters have default values. The second baseline was created with ResNet-50 model (row 2 of Table 1) which was trained using stochastic gradient descent with learning rate of 10^{-3}, momentum of 10^{-6}, decay parameter set as 0.9 and nestrov set to be true. Further, the networks presented in rows 3, 4, 5 of Table 1 were trained using Adam optimizer with learning rate of 10^{-4}, epsilon set to be 10^{-8} and remaining parameters set to default values. The CRNN model (row 3 of Table 1) comprises of concatenation of ResNet-50 and LSTM models while attention based sequence model (row 4 of Table 1) consists of augmentation of attention block over ResNet-50 model. The number of epochs for training Deep CNN, ResNet-50, CRNN, attention based model and Res-CRANN were set to be 150, 30, 100, 80 and 80, respectively. We observe validation results at each epoch and track model parameters corresponding to the lowest validation loss, which were later used for testing. All the networks were implemented using Theano[2] and Keras[8].

5.3. Results and Discussions

Table 1 shows the empirical results obtained using the proposed Res-CRANN architecture and baseline methods. We have evaluated percentage top-k classification accuracy with k set to 1, 3, and 5. The performance of a traditional

S.No.	Network Architecture	Top-1 Accuracy (%)	Top-3 Accuracy (%)	Top-5 Accuracy (%)
1.	Traditional Deep CNN	87.50	96.25	98.33
2.	ResNet-50	87.64	95.55	98.33
3.	CRNN	90.27	96.80	98.33
4.	Attention based Sequence model	90.27	96.25	97.92
5.	Res-CRANN (our method)	**90.42**	96.66	98.19
6.	Ensemble of (2), (3), (4) and (5)	**91.94**	**97.77**	**99.03**

Table 1. Table showing comparison of classification accuracy of our proposed network Res-CRANN with that of baseline deep networks for chromosome classification

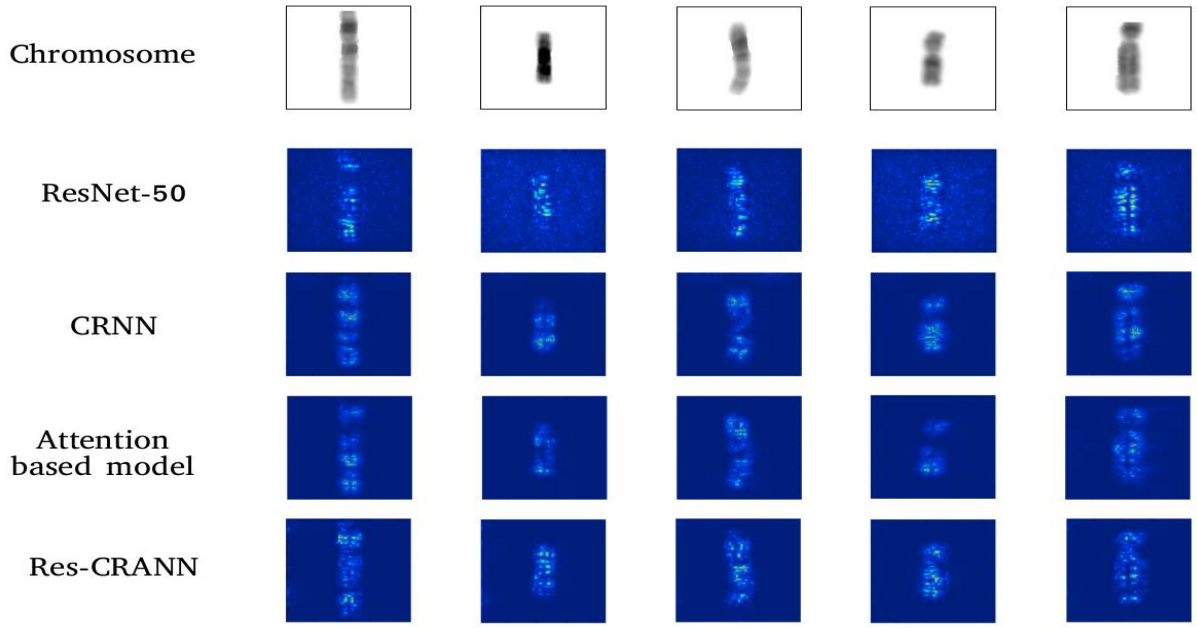

Figure 5. The examples of saliency maps produced by different networks. It shows that ResNet-50 network produces saliency maps with a lot of noise which means it does not localize chromosome features properly in the image. When we create saliency maps using augmented ResNet-50 network with recurrent layers and attention, i.e. Res-CRANN network, it produces a very fine and localized noise-less saliency map showing chromosome and corresponding sequence of bands as the region-of-interest.

Deep CNN model is shown in row 1 of Table 1. It comprises of 6 convolution layers with 16, 16, 32, 64, 128 and 256 as number of filters respectively, followed by 2 fully connected layers of dimensions 1024 and 512. Each convolutional layer uses Rectified Linear Units (ReLU) and is followed by a Max-pool layer of size 2×2. We have used 3×3 kernel sized filters in all the convolutional layers. The fully connected layers have *sigmoid* as their activation function. The last layer is also a fully connected softmax layer having 24 hidden units for 24 chromosome classes with a *softmax* activation function. As evident in row 1 and row 2 of Table 1, the improved performance of ResNet-50 model as compared to Deep CNN encourages us to use ResNet-50 in the CRNN architecture.

Subsequently, row 3 of Table 1 represents the performance of CRNN model which consists of ResNet-50 model in conjunction with LSTM as discussed in Section 3.2. This method gives the absolute improvement of **2.63%** and **2.77%** in Top-1 classification accuracy when compared with ResNet-50 and Deep CNN model, repsectively. This improvement demonstrates that using convolutional neural networks alone does not allow the learning of the sequence of chromosome bands completely. Thus, sequence information is explicitly incorporated using recurrent neural networks in CRNN model which improves the chromosome classification accuracy. Next, row 4 of Table 1 shows the results for Attention based Sequence model which is an augmentation of attention block to ResNet-50 model as discussed in Section 3.3, giving comparable performance to that of CRNN model while outperforming both baseline methods. This improvement is the result of learning the relationship between attentional region localization and classification.

In Table 1, row 5 shows the performance of our proposed method i.e., Res-CRANN model which achieves the Top-1 classification accuracy of **90.42%**, that outperforms previous state-of-the-art algorithms for chromosome classification.

Following this, we conduct another experiment in which we ensemble the results achieved from row 2, 3, 4 and 5 of Table 1 by taking the average of the probabilities obtained from these models. This gives us the Top-1 classification accuracy of **91.94%** which is an absolute improvement of **1.53%** over our Res-CRANN network. Additionally, top-3 and top-5 classification accuracies of Res-CRANN model are also highest when compared with the rest of the models. This shows that all of the three main layers, i.e., convolutional layers, recurrent layers and attention layers contribute to the improved performance of chromosome classification.

We also visualized saliency maps of some chromosomes produced by different network architectures, for example, ResNet-50, CRNN, Attention based CRNN model and Res-CRANN as shown in Figure 5. We used keras-vis [19] library for generating saliency maps. As we can see, ResNet-50 produces unwanted noise along with the chromosomal regions and hence, affects the classification accuracy of this network. However, when we augment ResNet-50 with recurrent layers and an attention mechanism (Res-CRANN), the network tries to pay attention only to the chromosome image regions, ignoring the background. This is made evident by the fine saliency maps, shown in Figure 5, in which only the area of interest, i.e. chromosome and sequence of bands is highlighted. This visualization demonstrates the improved performance of our architecture over other networks for chromosome classification.

6. Conclusion

This paper started with the introduction of the chromosome karyotyping process and explaining the motivation behind its automation. We proposed to utilize the sequence of dark and light colored bands present in the chromosome images in order to improve upon the classification accuracy of chromosomes as compared to baseline Deep CNN and ResNet-50 architecture. To achieve this, we proposed an end-to-end trainable Residual Convolutional Recurrent Attention Neural Network (Res-CRANN) which utilized the sequence of feature maps produced by convolutional layers of ResNet and subsequently, fed them to the recurrent neural networks. Further, we augmented the network with attention layers to make the network learn to concentrate on regions-of-interest. We observed via experimentation that Res-CRANN architecture proves to be robust and gives higher classification accuracy as compared to the state-of-the-art methods available for chromosome classification.

Acknowledgement

We would like to thank Professor Jeffrey D. Ullman from Stanford University for his valuable guidance and suggestions in improving the classification accuracy of chromosomes.

References

[1] Dataset online. `http://bioimlab.dei.unipd.it/Chromosome%20Data%20Set%204Seg.htm`.

[2] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, et al. Theano: A python framework for fast computation of mathematical expressions. *arXiv preprint*, 2016.

[3] J. Ba, V. Mnih, and K. Kavukcuoglu. Multiple object recognition with visual attention. *CoRR*, abs/1412.7755, 2014.

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

[5] A. P. Britto and G. Ravindran. A review of cytogenetics and its automation. *Journal of Medical Science*, 7:1–18, 2007.

[6] A. Carothers and J. Piper. Computer-aided classification of human chromosomes: a review. *Statistics and Computing*, 4(3):161–171, Sep 1994.

[7] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[8] F. Chollet. keras. `https://github.com/fchollet/keras`, 2015.

[9] J. K. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio. Attention-based models for speech recognition. In *Advances in Neural Information Processing Systems*, pages 577–585, 2015.

[10] D. Ciregan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE, 2012.

[11] P. A. Errington and J. Graham. Application of artificial neural networks to chromosome classification. *Cytometry Part A*, 14(6):627–639, 1993.

[12] A. Graves, A.-r. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*, pages 6645–6649. IEEE, 2013.

[13] K. Gregor, I. Danihelka, A. Graves, and D. Wierstra. DRAW: A recurrent neural network for image generation. *CoRR*, abs/1502.04623, 2015.

[14] 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*, pages 770–778, 2016.

[15] K. M. Hermann, T. Kocisky, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom. Teaching machines to read and comprehend. In *Advances in Neural Information Processing Systems*, pages 1693–1701, 2015.

[16] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[17] S. Jahani and S. K. Setarehdan. An automatic algorithm for identification and straightening images of curved human chromosomes. *Biomedical Engineering: Applications, Basis and Communications*, 24(06):503–511, 2012.

[18] S. Jindal, G. Gupta, M. Yadav, M. Sharma, and L. Vig. Siamese networks for chromosome classification. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[19] R. Kotikalapudi and contributors. keras-vis. `https://github.com/raghakot/keras-vis`, 2017.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[22] B. Lerner, M. Levinstein, B. Rosenberg, H. Guterman, L. Dinstein, and Y. Romem. Feature selection and chromosome classification using a multilayer perceptron neural network. In *IEEE International Conference on Neural Networks*, volume 6, pages 3540–3545, 1994.

[23] N. Madian and K. Jayanthi. Overlapped chromosome segmentation and separation of touching chromosome for automated chromosome classification. In *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5392–5395, 2012.

[24] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu. Recurrent models of visual attention. *CoRR*, abs/1406.6247, 2014.

[25] R. Pascanu, T. Mikolov, and Y. Bengio. Understanding the exploding gradient problem. *CoRR*, abs/1211.5063, 2012.

[26] J. Piper, E. Granum, D. Rutovitz, and H. Ruttledge. Automation of chromosome analysis. *Signal Processing*, 2(3):203–221, 1980.

[27] E. Poletti, E. Grisan, and A. Ruggeri. Automatic classification of chromosomes in q-band images. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1911–1914, Aug 2008.

[28] M. J. Roshtkhari and S. K. Setarehdan. A novel algorithm for straightening highly curved images of human chromosome. *Pattern Recognition Letters*, 29(9):1208 – 1217, 2008.

[29] M. Sharma, O. Saha, A. Sriraman, L. Vig, R. Hebbalaguppe, and S. Karande. Crowdsourcing for chromosome segmentation and deep classification. In *CVPR 2017*. IEEE CVPR 2017, 2017.

[30] B. Shi, X. Bai, and C. Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2016.

[31] S. Squartini, A. Hussain, and F. Piazza. Preprocessing based solution for the vanishing gradient problem in recurrent neural networks. In *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, volume 5, pages V–713–V–716 vol.5, May 2003.

[32] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[33] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2015.

[34] H. Xu and K. Saenko. Ask, attend and answer: Exploring question-guided spatial attention for visual question answering. In *European Conference on Computer Vision*, pages 451–466. Springer, 2016.