Deep Phenotypic Cell Classification using Capsule Neural Network

Subhankar Chattoraj[1], Arnab Chakraborty[2], Akash Gupta[3], Yash Vishwakarma[4]

Karan Vishwakarma[5], Jeetashree Aparajeeta[6]

[1,3,6]*Vellore Institute of Technology, Chennai, India.*

[2]*Chandigarh University, India.*

[4]*AKS University, Satna, MP, India.*

[5]*University of Paris-Saclay, France*

Abstract— **Recent developments in ultra-high-throughput microscopy have created a new generation of cell classification methodologies focused solely on image-based cell phenotypes. These image-based analyses enable morphological profiling and screening of thousands or even millions of single cells at a fraction of the cost. They have been shown to demonstrate the statistical significance required for understanding the role of cell heterogeneity in diverse biologists. However, these single-cell analysis techniques are slow and require expensive genetic/epigenetic analysis. This treatise proposes an innovative DL system based on the newly created capsule networks (CapsNet) architecture. The proposed deep CapsNet model employs "Capsules" for high-level feature abstraction relevant to the cell category. Experiments demonstrate that our proposed system can accurately classify different types of cells based on phenotypic label-free bright-field images with over 98.06% accuracy and that deep CapsNet models outperform CNN models in the prior art.**

I. INTRODUCTION

Image-based cell analytic methodologies enable understanding the cell heterogeneities and pattern development. Analyzing target cell images allows understanding the underlying spatial biological relation and determining the complex genetic and epigenetic-based workflows. The conventional methods of single-cell analysis are known as chemical cytometry [1] which is highly sensitive to analytical instruments. The chemical cytometer process and reports the data by quantifying one cell's cellular contents simultaneously is very expensive and time-consuming. The acquiring and analyzing of cell images are comparatively less complicated and economic processes. The tasks involving high latency and unprecedented processing throughputs such as cell screening and large-scale label-free rapid cell profiling [2] makes cellular image analysis a better alternative. However, the microscopy technologies currently exist scarifies the image fidelity by either producing high content images at low throughput or vice-versa. As a result, capturing the cell heterogeneity using image-based methodologies is still a significant challenge. Therefore, high-content cell images with high throughput are necessary for cell image analysis. Time-stretch microscopy, a new enhanced imaging technique, can help understand the cellular morphology and dynamics more accurately [3] [4] [5]. Bright field cell microscopic (BFCM) images obtained using Asymmetric-detection time-stretch optical microscopy (ATOM) imaging technique [3] can record ultra-fast, high contrast images of the human cells at high throughput. These images have proved their efficacy in detecting micro-particles in micro-fluid and can capture the morphological properties of the cells very accurately [6].

In a recent study, Fourier spectrum and wavelet coefficients have been used to extract useful information for cellular identification [7]. Textural-based features such as co-occurrence of adjacent local binary patterns (CoALBP) have also been utilized to learn the complex adjacent image micro-patterns distribution [8]. Recently, deep learning-based integrated techniques capable of learning directly from the training data have proven to outperform hand-crafted features in multidisciplinary domain [9] [10] and also eliminating the tedious process of selecting distinct feature extraction methods. The deep convolutional neural networks (CNNs) capable of extracting hierarchical and high-level features have manifested outstanding performance on generic visual-based recognition tasks such as in histopathological images classification [11] [12] and cell recognition and segmentation [13] [6]. Researches encompass rare events such as transition, cancerization, and differentiation for characterizing heterogeneous behavior at single resolution cell. The capturing of phenotypic changes of kinetic cells [14] and accurate identification of cell characteristics requires High-content screening (HCS) with high throughput, which relies on accurate fluorescent labels [15]. Fluorescent images are widely analyzed using CNNs models for learning staining patterns [16]. However, Fluorescence microscopy is restricted to a diffraction limit of $\sim 200nm$ for live-cell imaging, hence not capturing biological structures in a near-native state [17]. Bright-field images open new possibilities in understanding the biological cellular pattern. Recently, quantitative analysis of single living cells using deep CNN resulted in a reduction in curated time with high semantic segmentation accuracy [18]. In contrast, a computation method for predicting living cell division using dynamic features was reported in [19]. The alteration of morphology of colonies of embryonic stem cells using BFCM images was reported in [20] and the single-cell classification framework using (CNN) with to data size stability analysis was reported in [6]. However, it fails to utilize the spatial relations in biomedical imaging data due to its complex tissue structure [21]. Capsule Network [22] has been widely used in both segmentation and classification in medical imaging. Recently, SegCaps [23] was proposed for lung segmentation from CT scans via locally constrained routing and transformation matrix, while brain tumor classification using MRI images via capsule network was proposed in [21]. In this work, We have introduced a deeper modified Capsule Network for multi-class cell classification. Capsule networks are susceptible to various image backgrounds. Our proposed deep CapsNet is more reliable with changes in intensity variation in cellular images. The deep CapsNet
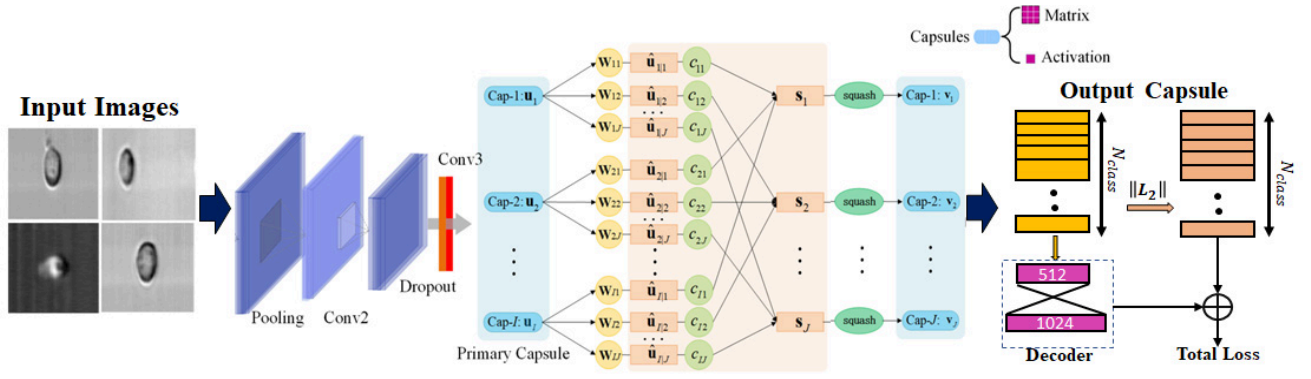
Fig. 1. Block diagram of the proposed Deep CapsNet architecture for automated human somatic BFCM images classification.

model works better on intensity variation cell images as the model trains on the principle of routing by agreement in which the capsules predict the outcome of the parent capsule at a lower level [21]. Accordingly, if the predictions agree, the parent capsules are activated. The proposed model is validated on more than 90,000 cell images to demonstrate its potency for deployment in clinical settings. At the same time, our experimental results outperform the existing state-of-the-art and several pre-trained-based models. Furthermore, our proposed deep CapsNet model eliminates the burden of manual segmentation and detailed annotations by the experts/radiologists. The main contributions can be recapitulated as:

- A Capsule neural network architecture for cell phenotypes classification framework, as shown in Fig. 1, is proposed. The model outperforms the state-of-the-art with better deep feature explainability.
- The Capsule framework leverages deep hierarchical feature space relationships among different cell phenotypes via deep instantiation parameters in a generalized discriminative manner.

Ongoing, the paper is structured as follows: Section II presents an overview of the previous works on cell classification. Section III describes the technical intricacies of the deep CapsNet architecture. Section IV discusses the experimental results, showing the comparative performance. Section V concludes the paper.

II. METHODOLOGY

A. Problem Formulation

The basic structure of CNN [24] comprises convolutional, pooling, and fully connected layers capable of extracting both higher-order local and elementary features. The computational cost is reduced significantly as the weights are shared over the whole input. The medical image processing [25] is benefited by the fact that these networks prior knowledge is non-essential to extract crucial features for accurate classification of class labels. In general, the CNN consists of K layers, the output $Y^{(p-1)}$ of layer $p-1$, for $(2 \leq p \leq K)$, the output Y^p is associated to the p input layer, which is given as:

$$X_{i,j}^{(p)} = \sum_{a=0}^{M-1} \sum_{b=0}^{M-1} W_{ab} Y_{i+a,j+b}^{(p-1)} \quad (1)$$

$$Y_{i,j}^{(p)} = \sigma\left(X_{i,j}^{(p)}\right) \quad (2)$$

The pre-activation output and size of the kernel are denoted by $X^{(p)}$ and M, respectively. The kernel matrix and the activation function are determined as W and $\sigma(\cdot)$ respectively. The CNN is also made translation-invariant by adding pooling or sub-sampling layer, reducing the trainable parameters. However, adding such layers makes it more difficult to compute the exact location of the feature descriptors, resulting in the network's incompetence in recognizing object-based images subject to different augmentation techniques [21]. Moreover, the images comprising different resolutions make it more difficult for the network to extract the underlying discriminative pattern. Our proposed deep CapsNet is invariant to augmentation and performs significantly with intensity variant images.

B. The proposed deep CapsNet Network

Deep CapsNet comprises a capsule with several neurons that encompass different instantiation parameters like size, rotation determining important features vector responsible for apprehending the probability of the presence of underlying objects [22]. The squashing non-linearity function is applied to construct a smaller length of such feature vector. The *Routing by Agreement* procedure is utilized instead of sub-sampling the feature maps of the pooling layers in the standard CNNs. The capsule contribution relies on how appropriately lower-level or child capsules predict the output of the consequent or parent capsules. The parent capsules account for the output of the child capsule only if an accurate prediction is provided. More particularly, the output and prediction of the lower-level i and higher-level j capsule can be denoted as u_i and $\hat{u}_{j|i}$ respectively. A linear transformation matrix W_{ij} is used to encode the relationship between i-th and j-th capsule which are propagated as: $\hat{u}_{j|i} = W_{ij}u_i$. The coupling coefficient c_{ij} or the strength determines the actual agreement between the predicted $\hat{u}_{j|i}$ and the actual output of s_j of the parent capsule j and the activation vector v_j of j is also computed as follows:

$$v_j = \frac{||s_j||^2}{1+||s_j||^2} \frac{s_j}{||s_j||}, \quad s_j = \sum c_{ij}\hat{u}_{j|i} \quad (3)$$

During the routing by agreement procedure, the log probability of whether the capsule i and j needs to be integrated or

not is signified as b_{ij}. The inner product a_{ij} between s_j and $\hat{u}_{j|i}$ vector is taken to compute the similarity for updating b_{ij} which is initialized to zero during the start of the training. Therefore a_{ij} is iteratively added to b_{ij} as follows:

$$a_{ij} = s_j . \hat{u}_{j|i} \tag{4}$$

while the softmax function is used to set the coupling coefficient c_{ij} as follows:

$$c_{ij} = \frac{exp(b_{ij})}{\sum_{k=1}^{K} exp(b_{ik})} \tag{5}$$

Where, the number of capsules is denoted as K. Each capsule j, for $(1 \leq j \leq K)$ a corresponding loss function is coupled with the last layer. The decoder is referred as fully connected layer which utilizes the the final instantiation parameters of the true classes as inputs to capture the hard to discern real representative features.

In comparison to the capsule loss, decoder contributes to the final error of the model with smaller weight. We utilized the margin loss proposed in [22] so that the top-level capsule k have a larger length only if $k(k \in \{0, 1\})$. The sum of the losses is the total margin loss L_j which is computed as:

$$L_j = \sum_k \left[\begin{array}{c} T_j max\left(0, m^+ - ||v_j||\right)^2 \\ +\lambda \left(1 - T_j\right) max\left(0, ||v_j|| - m^-\right)^2 \end{array} \right] \tag{6}$$

Where, $T_j = 0$ where class j is not present and $T_j = 1$ otherwise. The hyper parameter are initialized before the learning process and are denoted as $m^+ = 0.4$, $m^- = 0.2$ and $\lambda = 0.7$. The reconstruction loss L_r and the total loss L_{Total} is computed by the sum of squared differences between the input and reconstructed images as [26]:

$$L_{Total} = L_j + \alpha L_r \tag{7}$$

where, the $\alpha = 0.0005$ to minimize the reconstruction loss and scale down the margin loss during the learning phase.

III. EXPERIMENTAL RESULTS

A. Dataset Description

The proposed deep CapsNet model was validated on an open-source dataset available from [1]. The images were captured using ATOM imaging which provided significant-good cellular resolution, label-free, and high-contrast flow imaging. The dataset mainly comprises four classes MCF7, THP1, PBMC, and MB231. deep CapsNet model was trained on more than 1,00,000 images. The cell images were captured in different resolutions and focus at a very high speed to enhance the classification complexity of the images. The representative bright field cell microscopic images for all four classes are given in Fig 2. The resolution of cell images vary from 21×102 to 68×440 pixels. The primary objective is to model a robust classification model capable of classifying these cell images that are not sensitive to the representations of non-identical cell types images, capable of performing with higher accuracy and reliability. More details about the dataset can be found in [6].

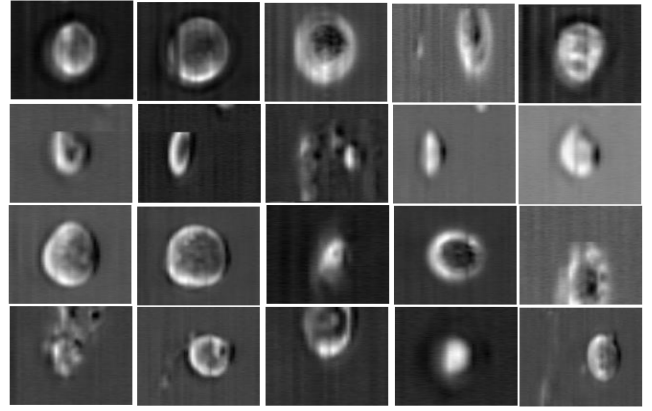[1] https://ieee-dataport.org/documents/human-somatic-label-free-bright-field-cell-images

Fig. 2. Representative Bright field cell microscopic images. Each row represents a specific type, from top to bottom: MCF7, THP1, PBMC and MB231

B. Experimental Settings & Performance Evaluation Metrics

The proposed deep CapsNet has been implemented using Tensorflow and Keras libraries in Python (Version: 3.5.2) programming environment. The hardware system comprises AMD Ryzen 7 3700× CPU (8 Cores/16 Threads), 64 GB of DDR4 RAM, NVIDIA RTX 3080 10GB VRAM graphics processing unit (GPU) while the software system includes Ubuntu 20.04 LTS ×64 OS with CUDA-enabled GPU as the parallel computing platform. The BFCM image dataset is split in the ratio 60%:20%:20% for training, validation, and testing. The distribution of cell images for each category was made uniform using the proper image augmentation technique to eliminate the class imbalance problem during the training phase. The proposed deep CapsNet network architecture can be traced in Fig 3. We utilized the Adam optimizer [27] with a learning rate and batch size of 10^{-3} and 32 respectively to ensure faster convergence during the training. A dropout rate [28] of 5% and L_2-regularization [29] is used to prevent model over-fitting. The evaluation metrics for cell image classification are divided into true-positive (TP), false-positive (FP), and false-negative (FN), where TP and FP determine the positive class cardinality of correct and incorrect predictions. In contrast, the FN determines the negative class. The multi-class classification assessment of deep CapsNet is tested based on the TP, FP, and FN values by the following performance evaluation metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{11}$$

The macro average is used to compute the overall precision, recall and F_1 score ensuring that deep CapsNet performs well both on common and rare classes of the BFMC images.

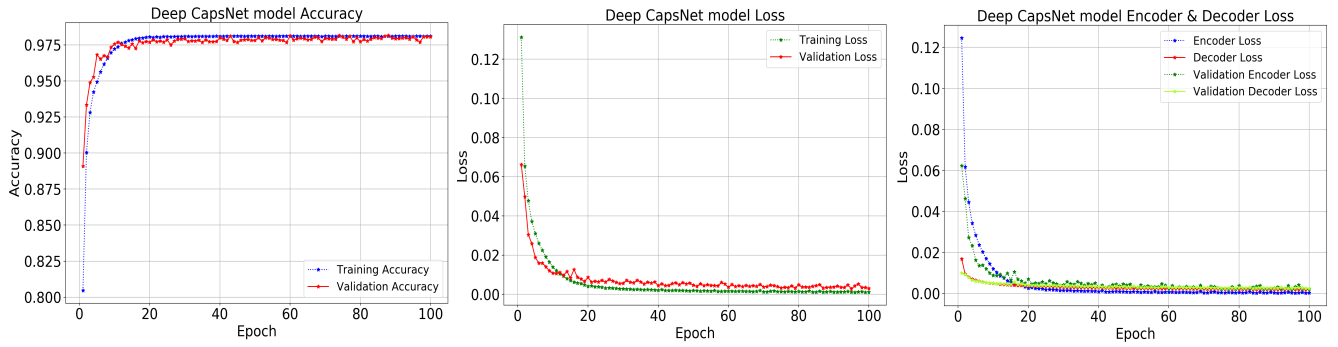$$Macro\_P = \frac{\sum_{i=1}^{c} Precision_i}{c} \tag{12}$$

Fig. 3. From left variation of training and validation accuracy, total loss, and encoder and decoder loss of deep CapsNet model over the epochs during network training.
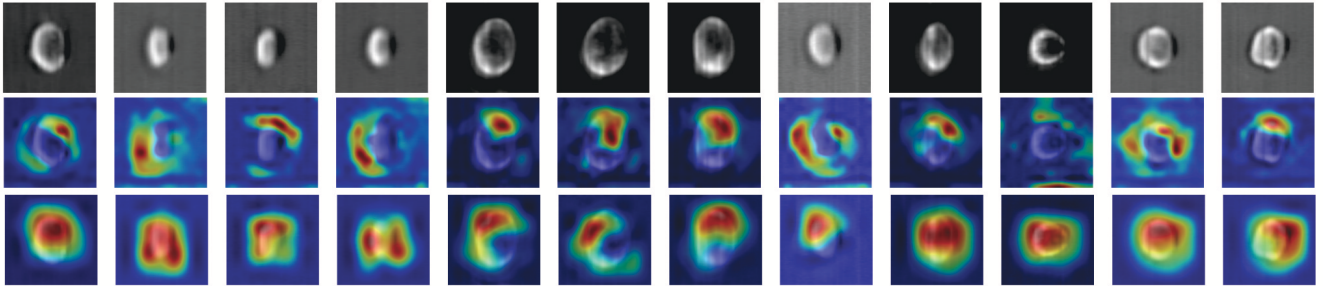

Fig. 4. Visualizing sample images (1st row) from the BFCM input image dataset, corresponding Grad-CAM localization of the ROI with the ResNet152V2 model (2nd row), and deep CapsNet model (3rd row).

$$Macro\_R = \frac{\sum_{i=1}^{c} Recall_i}{c} \qquad (13)$$

$$Macro\_F1 = \frac{\sum_{i=1}^{c} Precision_i}{c} \qquad (14)$$

C. Classification Results with the Proposed Model

The proposed deep CapsNet model has been evaluated on more than 1,00,000 images BIHM images [30] and evaluated using the performance evaluation metrics defined in Eqs. (8)–(14). The deep CapsNet accuracy, encoder, and decoder loss, and overall loss with increasing epoch are shown in Fig 4. It is observed that during the training, regardless of minimum fluctuation during the ascent, the accuracy increases continuously in successive iterations. The test dataset is fed into the deep CapsNet framework after re-slicing and shuffling. The data point generates independent changes on the model during the update without being biased, reducing variance and over-fitting. The validation loss drops while the validation accuracy increases and attains stability after 20 epochs, approximately signifying high convergence of the proposed model. According to the evaluation metrics, the deep CapsNet's high performance exhibits the capability to extract representative linearly separable underlying biological patterns from the BHFM images. Table I provides more comprehensive results of the classification performance. The average of several experimental results is taken to compute the final indices. More quantitatively, the overall accuracy score **98.06%**, $Macro\_P = \mathbf{97.76\%}$, $Macro\_R = \mathbf{96.89\%}$, and $Macro\_F1 = \mathbf{97.31\%}$ is determined. Table I indicate the class-specific indices of the deep CapsNet in terms of categorization and model assessment.

TABLE I
CLASS-SPECIFIC EXPERIMENTAL RESULTS

Metrics	MCF7	PMBC	THP1	MB231
Precision	96.87	98.22	98.18	98.18
F_1-score	96.40	96.79	98.75	98.75
Recall	95.94	95.39	99.34	99.34

IV. DISCUSSION

A. Comparative Evaluation of the deep CapsNet

1) Comparison with Pre-trained Models: Transfer learning (TL) using pre-trained deep learning models have been widely used as state-of-the-art image classification techniques. To prove the efficacy of our proposed deep CapsNet, we compared our performance with the pre-trained deep learning models. We fine-tuned all the pre-trained models to adapt deep discriminative feature pertaining to our classification. The current ImageNet weights [31] trained on more than 1000 classes, mean squared error (MSE) loss, and stochastic gradient descent (SGD) optimizer [32] was used for training and validation purpose. Deep CapsNet captured vital cell heterogeneous features that reflect superior classification performance significantly with every successive iteration compared to other TL models. In Fig. 4 and Fig. 4, the validation accuracy and loss over the epoch are given. In Table II. the quantitative comparison of the performance rate of the proposed deep CapsNet with other TL models is given. From Table II, it can be seen that deep CapsNet surpasses the recognition rate of more than **13.04%** for ResNet50 while

with the nearest competitor of **2.34%** for ResNet152V2. Our proposed model beat all the existing TL models with comparable trainable parameters and reduced network depth.
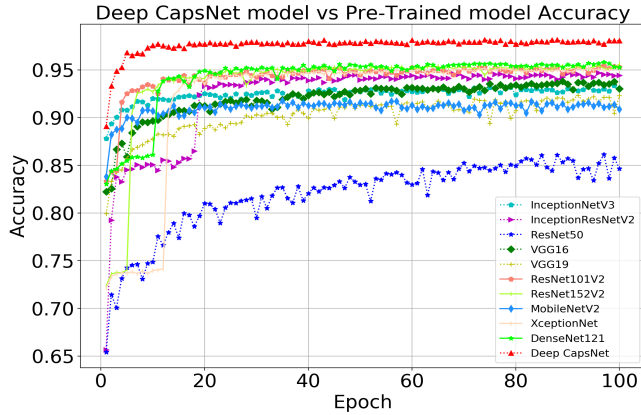

Fig. 5. Variation of accuracy in different pre-trained and deep CapsNet model over the epochs during network training.
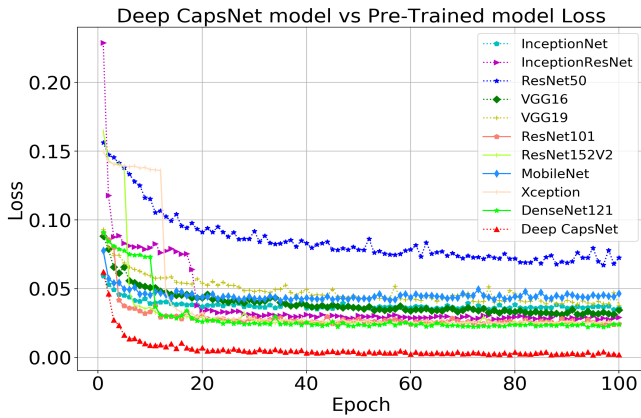

Fig. 6. Variation of loss in different pre-trained and deep CapsNet model over the epochs during network training.

TABLE II
COMPARATIVE EVALUATION OF PRETRAINED NETWORKS TRAINED ON BFCM DATASET.

Network	Depth	PM	Accuracy (%)
ResNet50	N.A	85.63	85.20
ResNet101V2	N.A	44.70	95.33
ResNet152V2	N.A	60.38	95.72
VGG16	23	138.3	93.39
VGG19	26	143.6	91.38
DesnseNet121	121	8	95.28
InceptionResNetV2	572	55.87	94.41
InceptionNetV3	159	23.85	93.19
Xception	126	22.91	95.13
MobileNetV2	88	3.53	91.19
This Work	**3**	**21.61**	**98.06**

B. Visual Explanations of Deep CapsNet Inference

The emergence of explainable artificial intelligence (XAI) provides a better understanding in decoding the black-box nature of DL framework's [33]. Models like deep CapsNet are trained on the whole images rather than labeled or extracted ROI. Recently proposed, gradient-weighted class activation mapping (Grad-CAM) aids in understanding the DL prediction with visual clarification [34]. The emergence of Grad-CAM, an attention visualization tool that assists in visualizing where a DL model is finding its ROI and understanding the decision of neurons based on the localization of the ROI. As illustrated in Fig 6. the deep CapsNet model recognizes the categories of BIHM whole-cell image pixels manifesting textural and cell boundaries. ResNet152V2 model is the nearest competitor to the proposed deep CapsNet model, which is also in agreement with the prior art [26]. As such, the ResNet152V2 model is chosen for understanding the visual explanation depicted by the Grad-CAM analysis. The Grad-CAM visualization generates the deep learning model's heat-map localization ROI in decision-making. The deep CapsNet exhibits better visual certainty in varying environments and more comprehendible than the ResNet152V2 model's heat-map ambiguities.

C. Analysis of the Network Performance

The ATOM imaging system is a unique imaging approach which opens new possibility in understanding complex biological pattern while achieving high-contrast flow label-free imaging with significantly better cellular resolution at a higher speed. This work is the first to envisage capsule network architecture in BHCM images with outperforming all TL models and present state-of-the-art to the best of our knowledge. The Grid search utilized in the same parameter space for hyper-parameter optimization efficiently optimizes the deep CapsNet model with improved learning consistency. The proposed architecture achieves high classification accuracy, making predictions based on whole-images rather than the ROI-based localization or segmentation approach in the conventional DL framework. The ability of the deep CapsNet network in reducing the high inter-class similarity and intra-class diversity in BHCM images demonstrates its efficacy from its predecessor. Henceforth, manifesting robust generalization incorrectly identifying images captured using ATOM-based imaging systems.

V. CONCLUSION

This article presents a deep Capsule network-based automated single-cell image classification system. The architecture comprises four convolutional layer kernels, PrimaryCaps layer, DigitCaps layer, and two fully connected layers and ReLU activations precede max-pooling layers with batch normalization operations in the network. We provide a concise overview of various aspects of this approach and address the impact of data volume on classification results, demonstrating that larger datasets support deep learning models. The suggested system outperformed other often-used approaches in experimental validation on large-scale dynamic datasets. With label-free photographs, the proposed method achieves over 98.06% accuracy in distinguishing different types of cells.

REFERENCES

[1] N. Dovichi and S. Hu, "Chemical cytometry." *Current opinion in chemical biology*, vol. 7 5, pp. 603–8, 2003.

[2] M. Boutros, F. Heigwer, and C. Laufer, "Microscopy-based high-content screening," *Cell*, vol. 163, pp. 1314–1325, 2015.

[3] T. T. W. Wong, A. K. S. Lau, K. K. Y. Ho, M. H. Tang, J. Robles, X. Wei, A. Chan, A. H. L. Tang, E. Lam, K. Wong, G. Chan, H. C. Shum, and K. Tsia, "Asymmetric-detection time-stretch optical microscopy (atom) for ultrafast high-contrast cellular imaging in flow," *Scientific Reports*, vol. 4, 2014.

[4] Q. T. K. Lai, K. C. M. Lee, A. H. L. Tang, K. Wong, H. K. H. So, and K. Tsia, "High-throughput time-stretch imaging flow cytometry for multi-class classification of phytoplankton." *Optics express*, vol. 24 25, pp. 28 170–28 184, 2016.

[5] J. Wu, Y. qing Xu, J. Xu, X. ming Wei, A. C. Chan, A. H. L. Tang, A. K. S. Lau, B. F. Chung, H. C. Shum, E. Lam, K. K. Wong, and K. Tsia, "Ultrafast laser-scanning time-stretch imaging at visible wavelengths," *Light, Science Applications*, vol. 6, 2017.

[6] N. Meng, E. Lam, K. Tsia, and H. So, "Large-scale multi-class image-based cell classification with deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, pp. 2091–2098, 2019.

[7] P. Soda and G. Iannello, "Aggregation of classifiers for staining pattern recognition in antinuclear autoantibodies analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, pp. 322–329, 2009.

[8] R. Nosaka, Y. Ohkawa, and K. Fukui, "Feature extraction based on co-occurrence of adjacent local binary patterns," in *PSIVT*, 2011.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[10] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137–1149, 2015.

[11] S. Pratiher and S. Chattoraj, "Diving deep onto discriminative ensemble of histological hashing class-specific manifold learning for multi-class breast carcinoma taxonomy," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1025–1029, 2019.

[12] S. Pratiher, S. Chattoraj, S. Agarwal, and S. Bhattacharya, "Grading tumor malignancy via deep bidirectional lstm on graph manifold encoded histopathological image," *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 674–681, 2018.

[13] M. Saha and C. Chakraborty, "Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation," *IEEE Transactions on Image Processing*, vol. 27, pp. 2189–2200, 2018.

[14] O. Z. Kraus and B. Frey, "Computer vision for high content screening," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 51, pp. 102 – 109, 2016.

[15] O. Z. Kraus, B. T. Grys, J. Ba, Y. Chong, B. Frey, C. Boone, and B. Andrews, "Automated analysis of high-content microscopy data with deep learning," *Molecular Systems Biology*, vol. 13, 2017.

[16] C. Kayasandik, W.-J. Ru, and D. Labate, "A multistep deep learning framework for the automated detection and segmentation of astrocytes in fluorescent images of brain tissue," *Scientific Reports*, vol. 10, 2020.

[22] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *NIPS*, 2017.

[17] R. Kaufmann, C. Hagen, and K. Grünewald, "Fluorescence cryo-microscopy: current challenges and prospects," *Current Opinion in Chemical Biology*, vol. 20, pp. 86 – 91, 2014.

[18] D. V. Valen, T. Kudo, K. Lane, D. N. Macklin, N. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. Ashley, and M. Covert, "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS Computational Biology*, vol. 12, 2016.

[19] A. R. Cohen, F. L. A. F. Gomes, B. Roysam, and M. Cayouette, "Computational prediction of neural progenitor cell fates," *Nature Methods*, vol. 7, pp. 213–218, 2010.

[20] N. Scherf, M. Herberg, K. Thierbach, T. Zerjatke, T. Kalkan, P. Humphreys, A. Smith, I. Glauche, and I. Roeder, "Imaging, quantification and visualization of spatio-temporal patterning in mesc colonies under different culture conditions," *Bioinformatics*, vol. 28, pp. i556 – i561, 2012.

[21] P. Afshar, K. Plataniotis, and A. Mohammadi, "Capsule networks for brain tumor classification based on mri images and coarse tumor boundaries," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1368–1372, 2019.

[23] R. LaLonde, Z. Xu, S. Jain, and U. Bagci, "Capsules for biomedical image segmentation," *Medical image analysis*, vol. 68, p. 101889, 2020.

[24] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," 1998.

[25] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. P. L. Lo, and G. Yang, "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 4–21, 2017.

[26] A. Mobiny, H. Lu, H. V. Nguyen, B. Roysam, and N. Varadarajan, "Automated classification of apoptosis in phase contrast microscopy using capsule network," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1–10, 2020.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[29] T. Van Laarhoven, "L2 regularization versus batch and weight normalization," *arXiv preprint arXiv:1706.05350*, 2017.

[30] "MS Windows NT kernel description," https://dx.doi.org/10.21227/H2QW97, accessed: 2020-08-12.

[31] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[32] S. Ruder, "An overview of gradient descent optimization algorithms," *ArXiv*, vol. abs/1609.04747, 2016.

[33] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "Xai—explainable artificial intelligence," *Science Robotics*, vol. 4, no. 37, 2019.

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.