Classifying Proteins using ResNet50 and InceptionV3

SWAMINI HEMANT BARKALE
Machine Learning / Deep Learning Engineer, R3 Systems India Private Limited

***Abstract*— Proteins are termed as the executing blocks of the human cell. Proteins serve several functions like structural support provision, protection against diseases, detoxing and regulation of the passage of materials across cell membrane, to name a few. Historically, the classification of proteins in human cell has been limited to a singular pattern in one or a few type of cells. In order to fully understand the complexity and working of the human cell, it is necessary to develop models that would classify mixed patterns across a wide range of human cells. This work provides a comparative study between two convolutional neural network models like ResNet50 and InceptionV3 for multi label classification of various proteins present in human cell. Handling of class imbalance is also tackled in the same.**

***Index Terms*— Convolutional neural network, InceptionV3, Multi label classification, ResNet50**

I. INTRODUCTION

Computer vision is a part of the deep learning technology that trains a model using image inputs to achieve a pre described goal. Computer vision technology has highest range of applications in areas of security and biological fields. In biological field, computer vision technology can be used to diagnose various diseases by providing magnetic resonance imaging and x-ray images as inputs along with diagnosed diseases as labels. These labels can be binary, multi class and multi label. Binary label classification can be explained with examples like classification of a tumor in brain magnetic resonance imaging image as malignant or not malignant. Also an x-ray image of a bone may help to diagnose if the bone is fractured or not. Retinal images of a diabetic person can help diagnose various levels of severity of diabetic retinopathy in that person. The levels of severity can be healthy retina, mild diabetic retinopathy, severe diabetic retinopathy and alarmingly severe diabetic retinopathy. This example can be described as multi class classification in computer vision. If there is a need to classify various components present in a single image into various classes, it gets classified as a type of multi label classification. In this paper we are going to use multi label classification for classifying mixed patterns of proteins in human cells.

For the same, we have selected an image dataset consisting of images of human cells along with the labels of the classified proteins.

II. LITERATURE REVIEW

A customized architecture with an adaptive concatenation of the pooling and buffering layers in the classification parts that is the fully connected parts was proposed, which outperformed the SOTA method of multi label protein classification by 2% in the F1 score [1].

An experiment was performed with the CSPNet algorithm on the original dataset before augmentation which provided a comparatively good result with respect to the existing research, which was 84% for the accuracy measure. By running a model for 9 epochs starting from 0 and ending at 8, it was observed that the accuracy increased from 94.2% at epoch 0 to 95% at epoch 8. ResNet Model was trained for 340 epochs and an accuracy of 91% was obtained. It was inferred that if the transferred learning approach would not been used, the accuracy obtained would not have been this better. Using BoTNet the accuracy reached was 93.8% in the second epoch. The idea for self-attention paved the way for increased accuracy [2].

Logical relationship and distribution characteristics among various labels were analyzed to determine and predict different sets of proteins present in different cells. Along with this comparison experiments were conducted on pre-trained Xception and InceptionResNet V2 to finally optimize the two models in terms of data augmentation, model structure and channel setting to name a few. It was inferred that the finely optimized InceptionResNet V2 model outperformed in the classification task [3].

It was suggested that multi label classification of proteins using the convolutional neural network model called GapNet-PL and using an activation function of parametric rectified linear unit instead of scaled exponential linear unit showed better performance for many classification metrics. The result showed that the GapNet-PL model with the parametric rectified linear unit as the activation function gave the F1 score of 0.541 and a recall of 0.473 [4].

III. RESEARCH METHODOLOGY

A structured methodology was followed to achieve the desired outcomes. Each step followed in this methodological approach has a specific significance. Eliminating any of these steps may not provide the desired outcome. Defining the problem helps set the goal of the research. Exploratory data analysis gives a clear picture of the data distribution which helps to define the further customized steps to be taken. Data pre-processing helps convert the data into a format that is compatible for the training process. Training and validation dataset creation helps in the model learning and evaluation respectively. During model training the model is educated to predict certain outcomes.

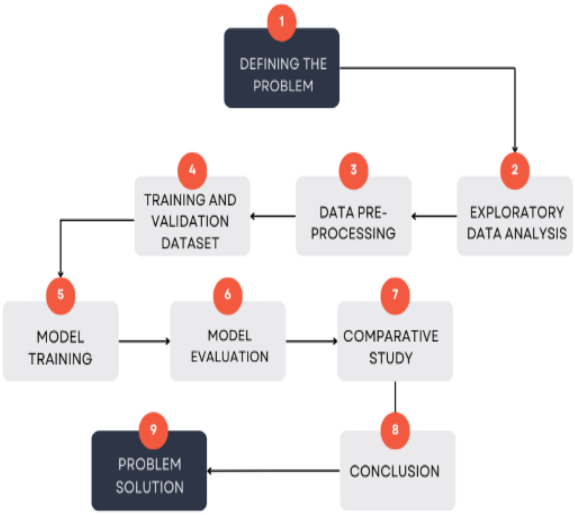

Fig. 1. Research Methodology

During model evaluation, the model is tested on whether the things it learned are correct or not. The comparative study compares the loss and accuracy obtained by the 2 models. Conclusion includes a systematic inference of the research. Problem Solution should typically be the goal of research.

1. Defining the problem: Due to lack of models to classify mixed patterns of proteins in microscopic images, it becomes difficult to fully understand the complexity and working of the human cell. The goal of this research is to develop a model that does multi label classification of mixed patterns of proteins in microscopic human cell images.

2. Exploratory data analysis: The dataset selected to achieve the goal was explored and studied. The entire dataset consisted of 124288 images consisting of 4 channels namely, red, green, blue, yellow for each image. Hence proper images to form a dataset were 31072. Each of these 31072 images had corresponding multiple labels describing mixed patterns of proteins present in that cell. However there is a high imbalance in these class labels. Some of these labels occur multiple times while some occur rarely. This phenomenon is prone to affect the model negatively. This is because the model learns more about a particular label and less about the others. Hence during validation there is a high probability that the model predicts the classes it has learnt more. In order to address this class imbalance, the classes that are underrepresented are given more weights. This solution is seen to perfectly handle class imbalance. Also this concept is applied only to training dataset, as validation dataset is only for model evaluation.

3. Data Pre-processing: The data is provided in the following format: a train folder and a train.csv file. The train folder consists of separate channel images of 31072 human cell. Hence total number of images present in this folder are 124288. Fig. 2. , Fig. 3. , Fig. 4. , Fig. 5. show a single cell image split into 4 channels, namely red, green , blue and yellow respectively.

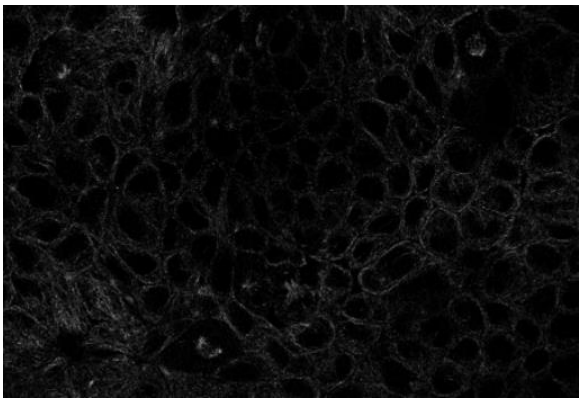

Fig. 2. Red Channel Image for a cell

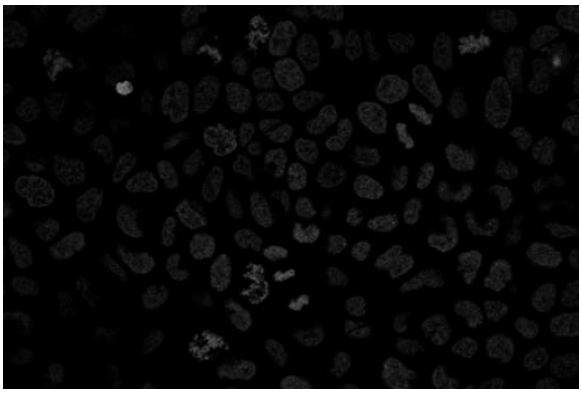

Fig. 3. Blue Channel Image for a cell

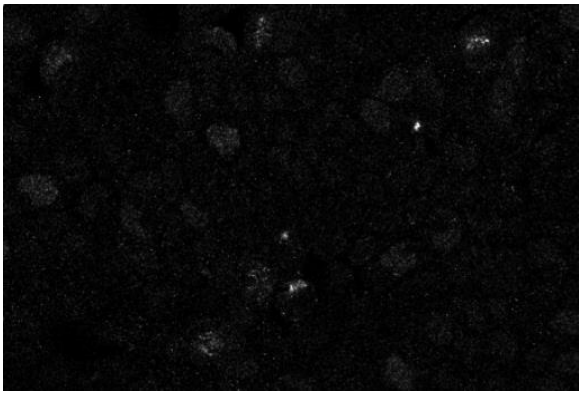

Fig. 3. Green Channel Image for a cell

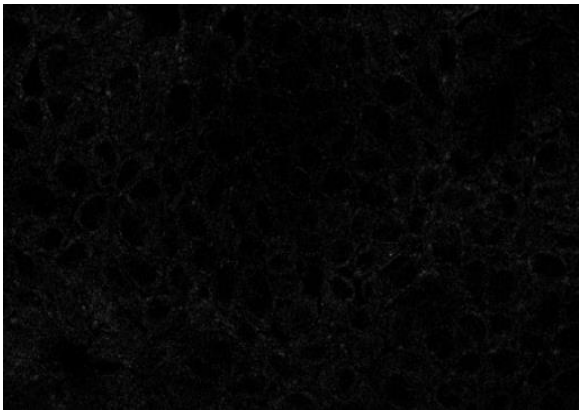

Fig. 4. Yellow Channel Image for a cell

It is necessary to stack all these 4 images together to see a single image consisting of all 4 channels. A class named HumanProteinClassificationDataset is defined to achieve this. Also the class mentioned above is so designed so that it returns image and its corresponding labels. This is necessary for further training of the model.

4. Training and Validation Dataset: It is necessary to split the dataset created into two parts. In this work, 95% of dataset is used for training and 5% for validation. Also for computational convenience, it is important to further create training and validation data loaders. Data augmentation can be achieved by creating data loaders. Data augmentation is not applied to validation data loader as the prime objective of validation dataset is only evaluation. Also shuffle property is set to false for validation data loader. As it is a comparative study, it is necessary to have the same data in the training and validation data loaders for every run.

Fig. 5. and Fig. 6. show batch of training and validation data loaders.

5. Model Training: Transfer learning approach is used to train the model. A comparison has been made between the loss and accuracy obtained by InceptionV3 and ResNet50 model. For InceptionV3 the size of the image required is 299 x 299 pixels. Hence, during application of transforms to both training and validation data loaders, the images are resized to 299 x 299 pixels. Models are trained in freeze state for 10 epochs. For first 5 epochs, a different learning rate is used as that for the next 5 epochs. Then in unfreeze state, the model is trained for 2 more epochs. OneCycleLR learning rate scheduler is used to approach the obtained accuracy.

6. Model Evaluation: The model is evaluated on the basis of the loss and accuracy obtained at the end of the 12^{th} epoch. Ideally, loss should decrease and accuracy should increase.

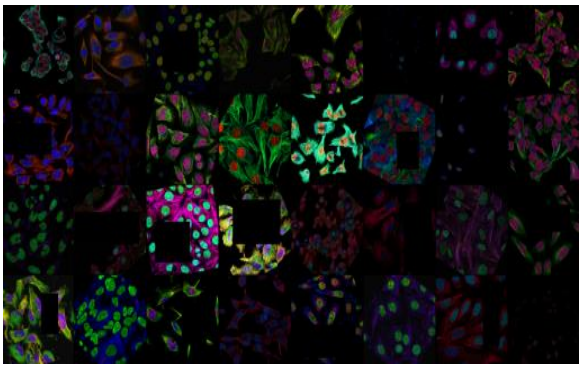

Fig. 5. A batch in training data loader

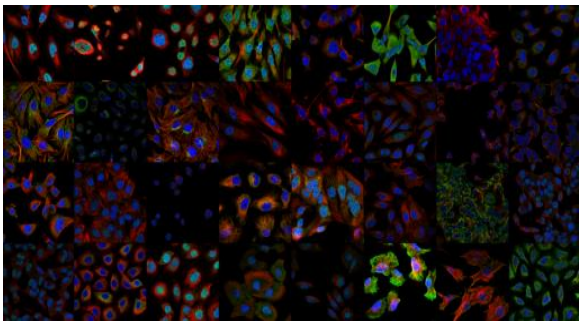

Fig. 6. A batch in validation data loader

7. Comparative study: The primary parameters selected to do the comparative study are loss and accuracy at the end of the 12th epoch.

8. Inference: By taking into consideration the 2 parameters mentioned above, a balanced inference is drawn.

9. Problem solution: If the obtained model or algorithm helps detect the mixed patterns of proteins in images from the real world, we have achieved a solution for the problem.

IV. FINDINGS

It was found that the ResNet50 model was better than InceptionV3 model for classifying mixed patterns of proteins in human cell. Though there was not a huge difference between the accuracies obtained by both the models, the loss proved to be the comparison and distinguishing factor. The loss obtained by the InceptionV3 method was decreased to 3.6901 and that obtained by the ResNet50 was decreased to 0.1688. Hence ResNet50 decreased the loss by a large extent as compared to InceptionV3.

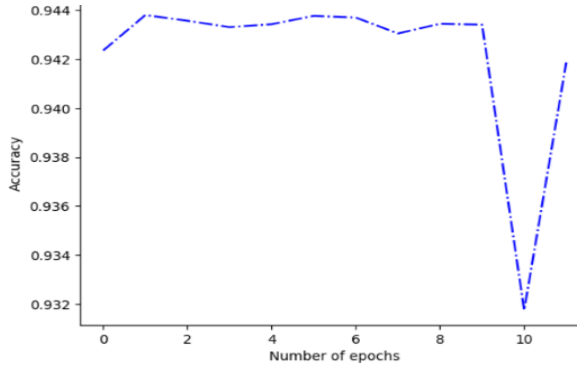

Fig. 7. Accuracy graph for ResNet50

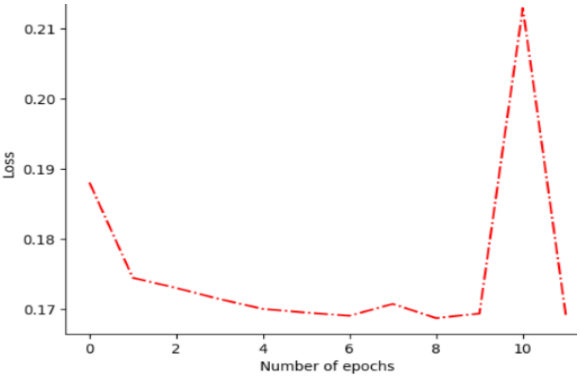

Fig. 8. Validation-Loss graph for ResNet50

Data augmentation that is application of transforms for the training data loader helped the model to learn better. This is because, data augmentation increases the size of the training data and reduces overfitting. Overfitting can be defined as, model performance is excellent on the data it is trained on but, poor on unseen data. Overfitting generally occurs when model tries to memorize the training dataset.

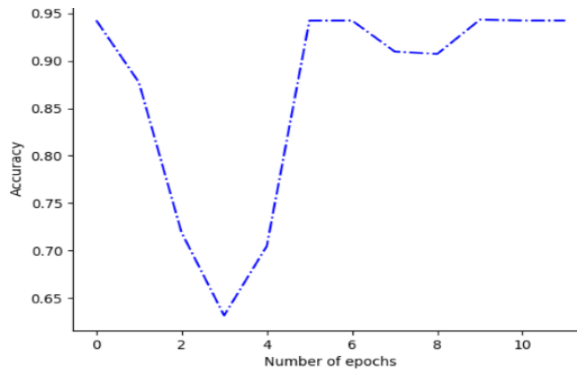

Fig. 9. Accuracy graph for InceptionV3

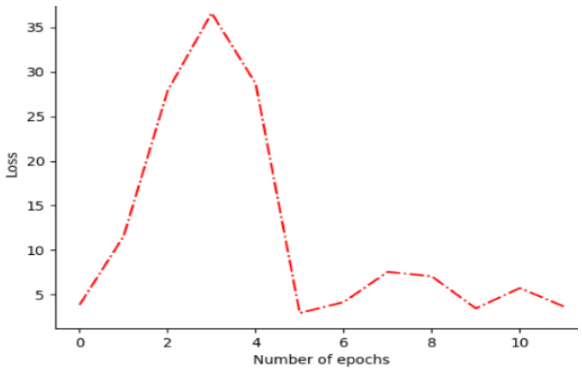

Fig. 10.Validation-Loss graph for InceptionV3

CONCLUSION

ResNet50 proved to be better than InceptionV3 due to the following reasons: ResNet50 is based on residual connections which makes the gradients flow more easily through the network during back propagation. This helps eliminate the vanishing gradient issue in deeper networks like the ResNet50, which enables the model to learn better and even complex representations. Moreover ResNet50 is excellent at capturing both low and high level features making it highly efficient to capture features in complex overlapping patterns like classifying mixed patterns of proteins in human cell. ResNet50 has 50 layers which is deeper than InceptionV3. This provides an additional advantage to ResNet50 model to catch very complex patterns in data. On the other hand, InceptionV3 uses multiple kernel sizes (kernel can be defined as a weight matrix) in parallel to capture features at multiple scale. This may not be effective for mixed patterns as in case of classifying mixed patterns of proteins. The loss was decreased to 0.1688 by ResNet50 and 3.6901 by InceptionV3. The accuracy obtained by ResNet50 and InceptionV3 respectively are 94.14% and 94.24%.

REFERENCES

[1] Enze Zhang, Boheng Zhang, Shaohan Hu, Fa Zhang, Zhiyong Liu, Xiaohua Wan. Multi-labelled proteins recognition for high-throughput microscopy images using deep convolutional neural networks, BMC Bioinformatics, 22, Article number:327

[2] Tahani Alsubait, Taghreed Sindi, Hosam Alhakami. Classification of the Human Protein Atlas Single Cell Using Deep Learning, applied sciences, MDPI

[3] Dong Yumin, Che Xuanxuan, Fu Yanying, Liu Hengrui, Sun Lina. Classification of human protein cell images using deep neural networks, IOS Press Content Library, Volume46, Issue2

[4] Lina Al-joudi, Muhammad Arif. Classification of Human Protein in Multiple Cells Microscopy Images Using CNN, Tech science Press.