A Cognitive Model for Classifying Human Sperm Morphology using Convolutional Neural Network

F. Josephine Nijofi Mactina[1], Dr. S. Neduncheliyan[2]

[1] *Research Scholar, Bharath Institute of Higher Education and Research, Chennai, India*
[2] *Professor/Dean, Dept. of Computer Science Engineering, Bharath Institute of Higher Education and Research, Chennai, India*
Email: [1]josephinemactina@gmail.com

Abstract: Evaluation of human sperm morphology is one of the important factors in the diagnosis of infertile male. Some Neural Network models was used to categorize the sperm morphology. In this proposed model, a deep neural network is drafted based on AlexNet, a pre-trained CNN. Human Sperm Head Morphology dataset (HuSHeM) was used in this model, which initially contains 216 RGB images. Training the model with 216 images was challenging and the accuracy obtained was oscillating between 65% to 69%. Data augmentation technique was used to overcome this challenge. The purpose to build the model is to automatically classify the sperm images based on the shape of the head and tail. Deep learning is put into the AlexNet in diverse ways such as, fine-tuning the weights of each layer and adding softmax layer, which has the cability to categorize the images accurately into the four different classes. Using this AlexNet, the training dataset obtained accuracy between 96% to 98% and testing dataset obtained 87% to 93%.

***Keywords*: AlexNet, HuSHeM, Deep Learning, Convolutional Neural Network, Data Augmentation**

1. Introduction

Deep Learning is a subset of Machine Learning, which is implemented successfully in various industries like healthcare, finance, media, retail and travel. Deep learning is otherwise known as hierarchical or deep structured learning, which yields better output with a massive amount of data. Over a short span of time, the availability and sophistication of Artificial Intelligence has exploded with array of technologies, tools and strategies especially in healthcare. DL offers the ability to analyze the data and precision never seen before. DL models like AlexNet, Imagenet and GAN become more accurate as they process huge amount of data, filter those data through cascade of multiple layers and form a pattern to produce perfect correlations and connections. Convolution Neural Network(CNN) models are very much efficient in image classifications and yields increased accuracy. CNN imitates some features of visual cortex for classifying images.

In this proposed work, human sperm morphology data set is being used. Sperm morphology is the size and structure of the sperm, one of the main factors in the semen examination. According to WHO (World Health Organization), normal sperm will have oval head and elongated tail. Abnormal sperms will have either head or tail defect. There are 3 types of sperm abnormalities: Tampered, Pyriform and Amorphous. The model is built to identify the four different images of the sperm.

2. Related Works

Soroush Javadi et al. [1] proposed a method for diagnosing male infertility using MHSMA data set which contains 1540 sperm figures from 235 patients with male factor infertility.In this model, they train a deep CNN on mini batches produced from the training set to save the checkpoint with minimal loss value on the validation set. Mini batches gradient descent method was employed to train the model. To encounters the problem of class imbalance and inadequate training samples, over sampling and data augmentation approach were employed. Obtained accuracies were 84.74%, 83.86%, 94.65% in acrosome, head and vacuole respectively.

E.Miahi et al. [2] proposed a method for diagnose male infertility using MHSMA data set. This model is trained by a search algorithm called genetic neural architecture search(GoNas), which operates as a meta controller that finds the constrained exploration space of plain CNN architecture to overcome the imbalance, low resolution and noisy data set, a novel method called GeNAS-WH is proposed. Accuracy of 92.66%, 77.33% & 77.66% is Vacuole, head & acrosome abnormality respectively was reached.

Emmanuel Lawrence Omonigho et al. [3] proposed a work which concentrates mainly on augmentation techniques to improve the accuracy on a limited data set. In this work DCNN (Deep Convolution Neural Network) with a modified Alexnet was used to train the model and categorize the Brest cancer mammography images into two category namely, benign and malignant tumors. This proposed model produced 95.70% accuracy.

Harsh Sharma et al. [4] suggested a model to extract the feature and to classify the images for faster examination by the medical practitioner. Data set used in the proposed model is chest X-Ray for detecting pneumonia from kaggle. This data set consists of 5863 images. In their proposed work four different models were used: 1) with augmentation, with dropout, 2) with augmentation, without dropout, 3) without

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

augmentation, with dropout, 4) without augmentation, without dropout. Accuracy for these four models were 0.9068, 0.8932, 0.7980 and 0.7498 respectively. This model also mentioned some techniques to be included in future work to avoid overfitting.

Kamel H. Rahouma et al. [5] proposed a model using ECG dataset to diagnose three different hear problems like Arrhythmia, Normal sinus rhythm and Congestive heart failure. To proceed with the work, DCNN algorithms such as AlexNet, GoogleNet and Adaptive Neuro Fuzzy Inference System Classification – FCM (ANFISC-FCM) is been used. Among these three models, ANFISC-FCM produced higher accuracy rate of 99%. Whereas, AlexNet and GoogleNet produces 94% and 96.6% accuracy respectively.

Abhishek Samanta et al. [6] proposed a work on classifying images of Diabetic Retinopathy(DR) using Transfer learning based CNN architecture. In this work they used 3050 training images and 419 validation images and the model was trained for 4 different classes and achieved an accuracy result of 0.8836 on validation set and 0.9809 on a training set.

Arshia Rehman et al. [7], suggested a model, Computer Aided Tumor Detection method, to detect the brain tumor accurately than the traditional Convolutional Neural Network. In this work, three studies (AlexNet, GoogleNet and VGGNet) were conducted to categorize different stages in brain tumors like, meningioma, glioma and pituitary. Data augmentation were also implemented to increase the dataset and reduce the problem of overfitting. Among the three models, VGGNet attained the higher accuracy of 98.69 in classification and detection.

Tulasi Krishna Sajja et al. [8], emerged a model to find lung cancer. This model is built using a modified GoogleNet, in which the dense layer is replaced with sparse layer to reduce the input channel and minimize the problem of overfitting. The model was trained using many CNN models but the proposed Network attained the higher accuracy of 99.03%.

Maayan Frid-Adar et al. [9], suggested a method, in which image augmentation was carried out in the dataset to enhance the classification performance (sensitivity and specificity). Classical data augmentation (CNN-AUG) and synthetic data augmentation were implemented and the results were compared. CNN-AUG yields 78.6% sensitivity and 88.4%specificity and synthetic data augmentation yields 85.7% sensitivity and 92.4% specificity.

3. System Design And Implementation

A. Network Design and Architecture

Convolution Neural Network Models

A Convolution Neural Network is a type of Deep Learning algorithm used in image processing and classification. It is designed specifically to process pixel data. CNN is composed of input layers, output layer and hidden layers. Hidden layers consists of convolution layers, ReLU layers, pooling layers and fully connected layers. The input image is provided with the convolution operation by the convolution layer, which passes the details to the next layer. Pooling layer chains the output of clumps of neurons into a single neuron in the subsequent layer. Every neuron in one layer is associated to every other neuron in the next layer using the fully connected layers.

AlexNet

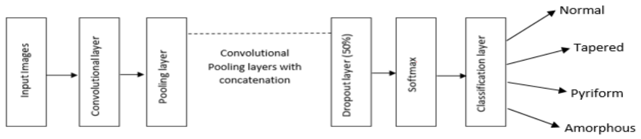


Fig. 1. Model architecture

Though CNN model can work efficiently with images and do not experience any overfitting problem, they are very much hard to process high resolution images in a short interval of time. The cost of training CNN model is high because of two main reasons. First is, the depth and the parameters of the images are huge in number. For example, AlexNet has 8 layers out of which there are 5 convolutional layers and 3 fully connected layers and it also has exceeding 60 million parameters. LeNet 5 has 7 layers. GoogleNet is composed of 22 layers and 6.8 million parameters. Second problem is processing massive image dataset of high resolution from ImageNet database.

AlexNet is composed of totally 8 layers out of which first 5 layers are the convolution layers and the last 5 layers are the fully connected layers. Each neuron in the convolution layer is associated only to the nearby neurons of the previous layer and all the neurons have same weight assigned. Pooling layer is added next to the convolutional layer, more specifically after ReLU layer, which can map the features given from the convolutional layer. This layer works on feature map independently and build a new set of features. Size of this layer is little than the size of the feature map. There are two regular functions applied in a pooling layer namely, average pooling, which calculates the standard value of the feature map and maximum pooling, which calculates the highest value of the feature map. Fully connected layer in contrast, every neuron in one layer is connected to every other neurons of the previous layer and each neuron are assigned different weight. This fully connected layer takes the result from the Convolutional Layer, fattens the result and convert it into a single vector for the next layer. The output from the last fully connected layer is being fed into the softmax function. Softmax function is used for the multi-task or multinomial classification. Mathematical representation for softmax function:

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

$$\boldsymbol{\sigma(z)_j = \frac{e^{\wedge} z_j}{\sum_{k=1}^{k} e^{\wedge} z_k}} \quad \text{for } j = 1,2,\ldots,k \qquad (1)$$

Rectified linear unit (ReLU) is an activation function and its mathematically defined as

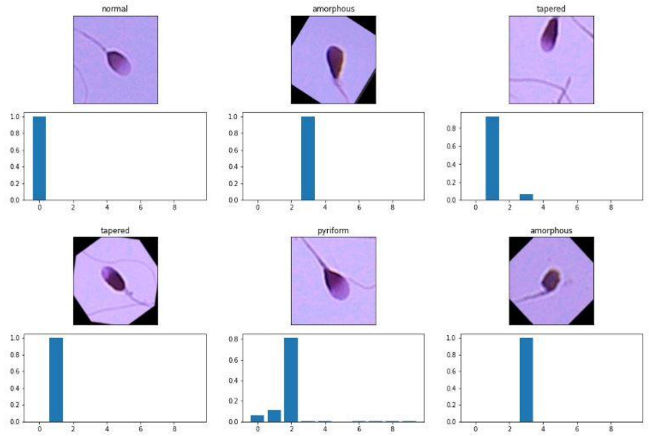
$$\boldsymbol{ReLU(x) = max(x, 0)} \qquad (2)$$

The gradient of ReLU is always 1, if the input is not less than 0. AlexNet has 60 million parameters and so arises a problem called overfitting. Overfitting develops when a designed model learns the nuances and noises in the trained data to an extent that it negatively impacts the performance of the model on new data, which leads to the poor performance of the designed model. This means that the noise or random fluctuations in the trained data is held up and learned as concepts by the trained model. The hitch is that these concepts do not apply to new data and negatively influence the model's ability to generalize. There are three ways discussed to reduce overfitting. First, simplify the model, which means reduce the complexity of the model by reducing the number of neurons or remove some layers to make the model smaller. Second, addition of dropout layers, which is pretty effective in reducing overfitting. Dropout layers actually drops some of the connection between layers. Third, early stopping, in this the training process must be stopped in the early stage, which means, instead of training the model for a fixed number of epochs, stop the training phase as soon as the validation loss rises.

B. Dataset Description

The dataset was built with the speculations prescribed by WHO 2010. The data is based on samples collected from patients aged between between 25 and 38 years, and semen smears were fixed and stained by Diff-Quik method. The Semen air dried initially and immersed in triarylmethane for 15 seconds, eosinophilic xanthene for 10 seconds, basophilic thiazine for 5 seconds and excess stain is removed under tap water for 10-15 seconds. The resolution of each images were 576×720 pixels in RGB color space. The sperm images were cropped and classified into four classes namely, normal, tapered, pyriform, and amorphous. Final dataset consists of the images of these four classes of sperm heads. The resulting dataset of sperm images which is denoted as Human Sperm Head Morphology dataset (HuSHeM) includes 216 sperm images (54 normal, 53 tapered, 57 pyriform, and 52 amorphous).

C. Augmentation



Performance of Deep Learning is often accelerated with huge amount of data, but in real time, availability of data is less in number. To overcome this problem, we use data augmentation. In this process the available images were transformed using several techniques like flipping, rotating, blurring, shifting and much more. Data augmentation is applied to the training dataset and not with the test or validation dataset. In this proposed model, HuSHem (Human Sperm Head Morphology) dataset were used. This dataset consists of 216 images, which is categorized into four, Normal, Tapered, Pyriform and Amorphous. There are several types of augmentation process [13], out of which geometric transformations were used in this model. Flipping: using this technique, the images were flipped both horizontally and vertically. Rotate: Rotation augmentations are done by rotating the image clock-wise and anti-clockwise to certain angle. Blur and Add noise: using this, the images are blurred to certain limit and noises were added to the images.

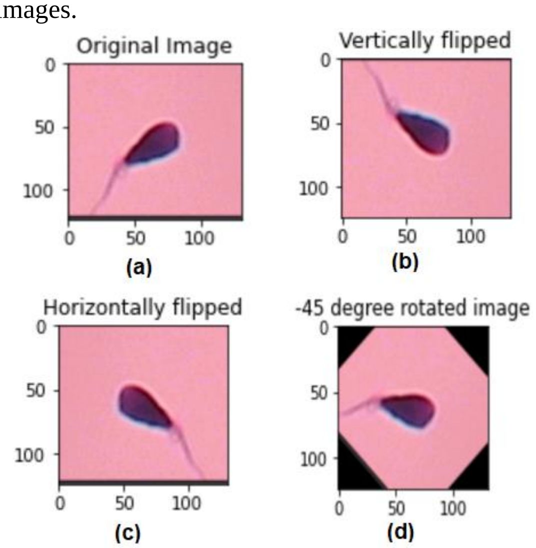


Fig.2 (a) Original image of the sperm dataset (b) original image vertically flipped (c) original image horizontally flipped (d) original image rotated 45 degrees anti-clock wise

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

4. Implementation And Results

A. Experimental Results

In this proposed work, four diagnosis classes were discussed: Normal, Tapered, Pyriform and Amorphous. As it was studied earlier, only limited studies were conducted on HuSHem dataset and anticipated using some CNN models for diagnosing the Sperm Morphology Analysis. The experimental analysis was carried out in google python colaboratory with necessary GPU. Initially the dataset was trained and classified using one of the CNN models, AlexNet with the original 216 datasets. The obtained accuracy was not feasible. To overcome such complexities, geometrical augmentation was carried out and the outcome was feasible and prediction was matched.

Figure 3 depicts the classification of the sperm image. Class 0 is Normal, class 1 is Tampered, class 2 is pyriform and class 3 is amorphous. The image dataset consists of 216, but later those images were augmented [13] for better classification. The whole dataset is split into train and test (70% and 30% respectively) dataset. After the process of augmentation, there were 2,706 images, with which the outcome of the model was improved. The model for trained for about 100 epochs on the RGB images. The accuracy varies between 0.87 to 0.93.

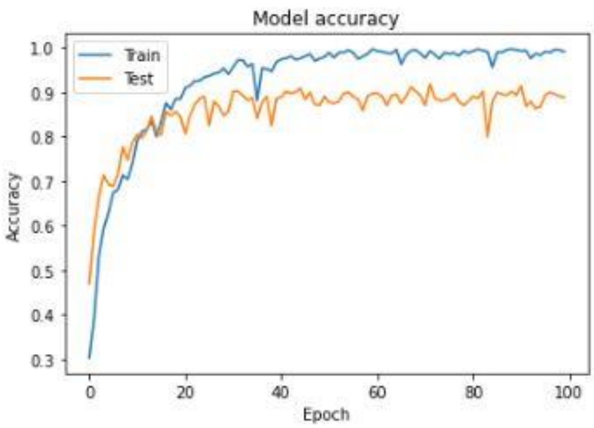


Fig.3 Accuracy graph for train and test data for 100 epochs

B. Performance Evaluation

Greater the effectiveness of the model, better the performance. To measure this performance in machine learning or deep learning, confusion matrix is implemented. The performance of the deep neural network classifier for the given dataset is evaluated based on precision and recall. These are the basic parameters of evaluation for the classification subsystem.

True Positive(TP):Abnormal instances correctlyidentified as abnormal.

True Negative(TN): Accurate negative forecast normal instances correctly identified as normal.

False Positive(FP): Inaccurate positive forecast abnormal instances incorrectly identified abnormal.

False Negative(FN): Inaccurate negative forecast normal instances incorrectly identified as abnormal.

Accuracy: Accuracy is defined as the sum of the correct positive prediction and the incorrect negative prediction divided by the sum of all true positive, true negative, false positive and false negative prediction.

Precision: The correct positive predictions divided by the total number of positive prediction as intended. Precision value is defined as the proportion of the true positives against all the positive results. A “true positive” is the event when the model makes a positive prediction, and the subject has a positive result and a “falsepositive” is the event when the model makes a positive prediction but the subject has a negative result.

Recall: Recall can be defined as the ratio of correct positive predictions to the sum of true positive and false negative prediction.

Sensitivity: Sensitivity is the measure of the proportion of positives that are correctly identified.

Specificity: Specificity is the measure of the proportion of negatives that are correctly identified. The above mentioned performance evaluation can be done using the following equations.

$$Accuracy = \sum_{i=1}^{l} \frac{(TP_i + TN_i)}{(TP_i + FN_i + TN_i + FP_i)} / l \quad (3)$$

$$Precision = \sum_{i=1}^{l} \frac{(TP_i)}{(TP_i + FP_i)} / l \quad (4)$$

$$Specificity = \sum_{i=1}^{l} \frac{(TN_i)}{(TN_i + FP_i)} / l \quad (5)$$

$$Recall = \sum_{i=1}^{l} \frac{(TP_i)}{(TP_i + FN_i)} / l \quad (6)$$

$$Sensitivity = \sum_{i=1}^{l} \frac{(TP_i)}{(TP_i + TN_i)} \quad (7)$$

5. Conclusion And Imminent Work

In this work, a deep learning model, Alexnet to build a CAD system to classify the sperm images among the four categories has been developed. To carry out the proposed system in an efficient way, huge datasets were needed, but the dataset used in this work contains very few images, which is most challenging part of this work. To overcome this issue, data augmentation was implemented to expand the dataset. The proposed model has no issues in detecting the different categories. The accuracy obtained from the proposed model yields less, so the same data set can be tested with different neural network model.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

References

[1] Soroush Javadi, Seyed Abolghasem Mirroshandel, A novel deep learning method for automatic assessment of human sperm images, computers in Biology and Medicine 109 (2019) 182-194J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] E. Miahi, S. A. Mirroshandel*, A. Nasr, Genetic Neural Architecture Search for automatic assessment of human sperm images, arXiv:1909.09432v1 [cs.LG] 20 Sep 2019.

[3] Emmanuel Lawrence Omonigho, Micheal David, Achonu Adejo, Saliyu Aliyu, Breast Cancer:Tumor Detection in Mammogram Images Using Modified AlexNet Deep Convolution Neural Network, 2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS) 978-1-7281-3126-9/20/$31.00 ©2020 IEEE 10.1109/ICMCECS47690.2020.240870.

[4] Harsh Sharma, Jai Sethia Jain, Priti Bansal, Sumit Gupta, Feature Extraction and Classification of Chest X-Ray Images Using CNN to Detect Pneumonia, 978-1-7281-2791-0/20/$31.00 c 2020 IEEE.

[5] Kamel H. Rahouma, Rabab Hamed M. Aly and Hesham F. A. Hamed, Applying Deep Learning Techniques for Heart Big Data Diagnosis, © Springer Nature Singapore Pte Ltd. 2020.

[6] Abhishek Samanta, Aheli Saha, Suresh Chandra Satapathy,∗, Steven Lawrence Fernandes, Yo-Dong Zhang, Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset, 0167-8655/© 2020 Elsevier.

[7] Arshia Rehman, Saeeda Naz, Muhammad Imran Razzak, Faiza Akram, Muhammad Imran, A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning, Imran's work is supported by the Deanship of Scientific Research, King Saud University through research group Project Number RG-1435-051.

[8] Tulasi Krishna Sajja, Retz Mahima Devarapalli, Hemantha Kumar Kalluri, Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning, Journal homepage: http://iieta.org/journals/ts.

[9] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, Hayit Greenspan, GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification, Preprint submitted to Neurocomputing.

[10] Mohammad Havaeia, Axel Davyb, David Warde-Farleyc, Antoine Biardc,d, Aaron Courvillec, Yoshua Bengioc, Chris Pal, Pierre-Marc Jodoina, Hugo Larochelle, Brain Tumor Segmentation with Deep Neural Network

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, http://lmb.informatik.uni-freiburg.de/.

[12] Barıs Kayalıbay, Grady Jensen, Patrick van der Smagt, CNN-based Segmentation of Medical Imaging Data.

[13] Tri-Cong Pham, Chi-Mai Luong, Muriel Visani, Van-Dung Hoan, Deep CNN and Data Augmentation for Skin Lesion Classification, Springer International Publishing AG, part of Springer Nature 2018

[14] Yan Xu, Zhipeng Jia, Yuqing Ai, Fang Zhang, Maode Lai, Eric I-Chao Chang, Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation, 947978-1-4673-6997-8/15/$31.00 ©2015 IEEE

[15] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Member, Tarek M. Taha, Vijayan K. Asari, Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation.

[16] HamzaO Ilhan, Onur Sigirci, Gorkem Serbes, Nizamettin Aydin, A fully automated hybrid human sperm detection and classification system based on mobile-net and the performance comparison with conventional methods, https://doi.org/10.1007/s11517-019-02101-y.

[17] Nour Eldeen M. Khalifa, Florentin Smarandache, Mohamed Loey, A Study of the Neutrosophic Set Significance on Deep Transfer Learning Models: An Experimental Case on a Limited COVID-19 Chest X-Ray Dataset, nourmahmoud@cu.edu.eg

[18] Anandhavalli Muniasamy, Sehrish Tabassam, Mohammad A. Hussain, Habeeba Sultana, Vasanthi Muniasamy, and Roheet Bhatnagar, Deep Learning for Predictive Analytics in Healthcare, Springer Nature Switzerland AG 2020 A. E. Hassanien et al. (Eds.): AMLTA 2019, AISC 921, pp. 32–42, 2020

[19] Abhishek Samanta a, Aheli Saha a, Suresh Chandra Satapathy a, Steven Lawrence Fernandes b, Yo-Dong Zhan, Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset, https://doi.org/10.1016/j.patrec.2020.04.026

[20] Connor Shorten, Taghi M. Khoshgoftaar, A survey on Image Data Augmentation for Deep Learning, Springer Publications.

[21] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L. Rubin, Bradley J. Erickson, Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions, Springer Publications.

[22] Min Chen, Xiaobo Shi , Yin Zhang, Di Wu, Mohsen Guizani, Deep feature learning for medical image analysis with convolutional autoencoder neural network, 2332-7790 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

[23] Feng Jiang, Aleksei Grigorev, Seungmin Rho, Zhihong Tian, YunSheng Fu, Worku Jifara, Khan Adil, Shaohui Liu, Medical image semantic segmentation based on deep learning, Springer publications.

[24] Adnan Qayyum, Syed Muhammad Anwar, Muhammad Awais, Muhammad Maji, Medical image retrival using deep convolutional neral network, www.elsevier.com/locate/neucom.

[25] Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, Ronald M. Summer, Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731