Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09304-8

Detection of Mesangial hypercellularity of MEST-C score in immunoglobulin A-nephropathy using deep convolutional neural network

Shikha Purwar[1] · **Rajiv Tripathi**[1] · **Adarsh Wamanrao Barwad**[2] · **A. K. Dinda**[2]

Received: 20 September 2019 / Revised: 19 June 2020 / Accepted: 6 July 2020 /
Published online: 28 July 2020

Abstract

Immunoglobulin A (IgA)-nephropathy (IgAN) is one of the major reasons for renal failure. It provides vital clues to estimate the stage and the proliferation rate of end-stage kidney disease. IgA stage can be estimated with the help of MEST-C score. The manual estimation of MEST-C score from whole slide kidney images is a very tedious and difficult task. This study uses some Convolutional neural networks (CNNs) related models to detect mesangial hypercellularity (M score) in MEST-C. CNN learns the features directly from image data without the requirement of analytical data. CNN is trained efficiently when image data size is large enough for a particular class. In the case of smaller data size, transfer learning can be used efficiently in which CNN is pre-trained on some general images and then on subject images. Since the data set size is small, time spent in collecting large data set is saved. The training time of transfer learning is also reduced because the model is already pre-trained. This research work aims at the detection of mesangial hypercellularity from biopsy images with small data size by utilizing the transfer learning. The dataset used in this research work consists of 138 individual glomerulus ($\times$ 20 magnification digital biopsy) images of IgA patients received from All India Institute of Medical Science, Delhi. Here, machine learning (k-nearest neighbour (KNN) and support vector machine (SVM)) classifiers are compared to transfer learning CNN methods. The deep extracted image features are used by machine learning classifiers. The different evaluation parameters have been used for comparing the predictions of basic classifiers to the deep learning model. The research work concludes that the transfer learning deep CNN method can improve the detection of mesangial hypercellularity as compare to KNN, SVM methods when using the small data set. This model could help the pathologists to understand the stages of kidney failure.

Keywords Artificial neural network · Convolution neural network · MEST-C score · Mesangial hypercellularity · Immunoglobulin A-nephropathy

✉ Shikha Purwar
shikhapurwar@nitdelhi.ac.in

Extended author information available on the last page of the article.

1 Introduction

Immunoglobulin A (IgA)-nephropathy (IgAN) is also known as Berger's disease in which the basic filtration unit of the kidney i.e. nephrons stop functioning efficiently. This happens because of the deposit of IgA antibody in glomerulus [9]. The glomerulus is located inside the Bowman's capsule which is situated in the renal corpuscle of the nephron [26]. Blood flows through the afferent arteriole (red), enters the renal cell and is filtered into the glomerulus then the blood exits the efferent artery (blue) as shown in Fig. 1. The filtrate goes out into the renal tubule. In the case of kidney is inefficient of filtering the blood, it shows some symptoms like tea-coloured urine, high blood pressure, and swelling in hands and feet. The main cause of tea coloured urine is the leakage of blood and protein in it. As of now, IgAN is incurable although medication exists that can slowdown its progress. IgAN in long term, if it is left untreated, it causes kidney failure. The disease is found commonly in people of Caucasian, Asian origin, and rarely in people of African origin.

IgA deposited inside the glomerulus are diagnosed by performing immunoflurocense process on the glomerulus images [9]. After the diagnosis of IgA, MEST-C scoring is done to predict its stage [8, 14]. Figure 2 shows the semantic of renal corpuscle [53]. Approximately 5-6 images of glomeruli are analyzed for each kidney biopsy using a microscope. For each glomeruli, M, E, S, T and C abnormalities are identified and scoring is done thereafter on the following basis:

1. Mesangial hypercellularity (M):- This refers to the proliferation of pink-colored mesangial cell which is located in between tuft of capillaries inside the glomerulus. This is defined as the presence of more than four mesangial cells in any mesangial area of the

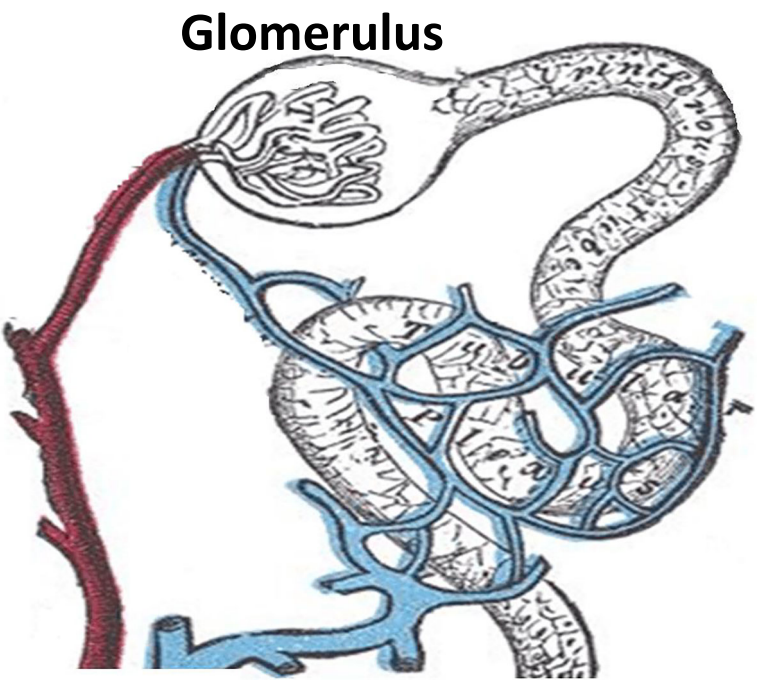


Fig. 1 The renal corpuscle [26]

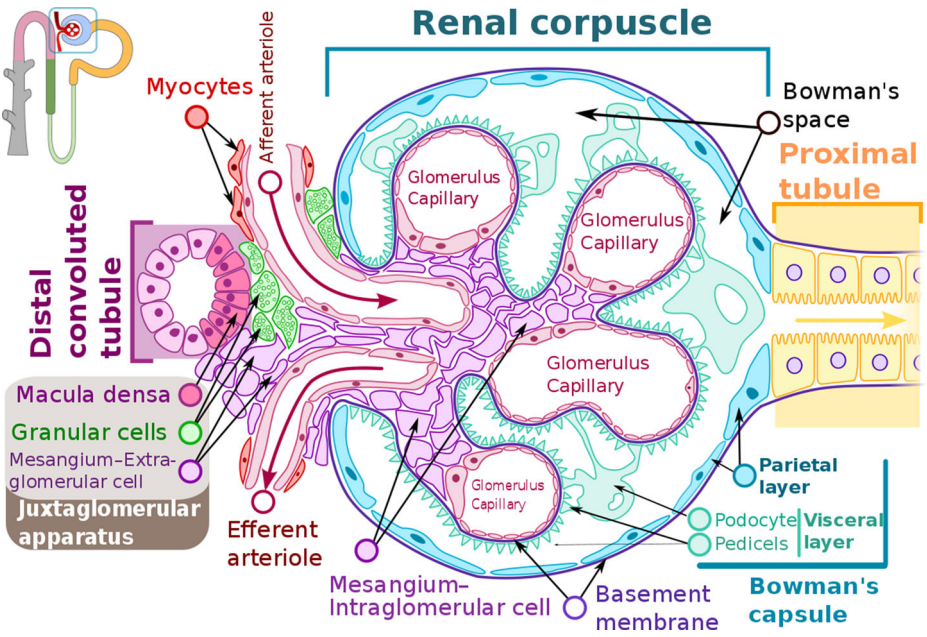


Fig. 2 Renal corpuscle internal structure [53]

glomerulus. If this is present in less than 50% of the glomeruli, the score is given as zero. Score 1 is given in the case of more than 50% of the glomeruli are affected.

2. Endocapillary hypercellularity (E):- This refers to the proliferation of endothelial cells in the lumen of the capillary leading to narrowing of capillary passage for movement of blood cells. A score 0 is given if no endocapillary hypercellularity is present, while a score is given as 1 if the same is present in any of the glomeruli.
3. Segmental Sclerosis (S):- This is defined as scarring in less than half of the glomerular tuft. Score 0 is given when sclerosis is absent and 1 is given when it is present in any of the glomeruli.
4. Tubular Atrophy (T):- This refers to the thickening of basement membranes inside the tubules resulting in the reduction of overall tubular diameter. It is scored on the basis of the percentage of the cortical area involved. Score 0 is given in case 0-25% of the cortical area is occupied, 1 is given in case 26-50% of the cortical area is occupied, 2 is given in case more than 50% of the cortical area is occupied.
5. Crescents (C):- This is defined as spreading, of extracapillary cells in the Bowman's space. Presence of usually at least two layers of these cells in the circumference of Bowman's space is termed as a crescent. Score 0 is given if it is absent, 1 is given if it is present in 0-25% of glomeruli, 2 is given if it is present in more than 25% of the glomeruli.

The staining process is done after images obtained from kidney biopsy. The pathologist analyzes these stained images using different magnification and resolution factors. The pattern, texture and other morphological features of the biopsy image are studied during the microscopic examination. In order to study these characteristics, continuous panning, zooming, and focusing on biopsy image under the microscope is required. This process requires

very high concentration and is therefore very tiresome and time-consuming. The exhaustive nature of the process increases the probability of manual error which sometimes leads to misdiagnoses. As a solution to be problem explained above, there has been the advancement of digital imaging techniques combined with machine learning to automate the process of analyzing the pathological images and its morphological features.

There have been several machine learning approaches like support vector machine (SVM), k-nearest neighbor (KNN), and random forest (RF) used to classify the disease based on the features. These features are extracted by several feature extractor such as scale invariant feature extraction (SIFT), local binary patterning (LBP), histogram of oriented gradients (HOG), and speeded up robust features (SURF). Although SVM is a black-box process, the parameters of SVM can be fine-tuned with the help of appropriate visualization tools[34]. These approaches have achieved success in the diagnoses of several clinical conditions such as breast cancer [36], Alzheimer's Disease [18], epilepsy [28], cardiovascular diseases [7], diabetic retinopathy [45], cytomegalovirus disease [47], liver transplantation [16], blood pressure estimation [5, 6], Parkinson [37], celiac disease [52], and multiple sclerosis diseases [19] with the help of good feature extractor and a sturdy classifier.

The literature of IgAN and other renal conditions has been explained here. Statistical techniques, such as the uni-variate and multiple linear regression model have been used to predict IgAN using data of the patients already diagnosed with the disease through biopsy procedure [3, 9, 11, 25, 35, 43, 44].

Geddes et al. [24] trained artificial neural network (ANN) on routine clinical information of a fixed set of 54 people to classify the kidney condition as 'stable' and 'non-stable'. After 7 years, 'stable' condition is referred to as the presence of less than 150 μmol/l serum creatinine and 'non-stable' condition as presence of greater than 150 μmol/l serum creatinine. Here, the outputs of ANN and six experienced nephrologists are compared and provided 87.0% accuracy, 86.4% sensitivity or TPR, and 87.5% specificity or TNR.

Cannone et al. [13] used classification tools like naive bayes, SVM, decision tree, and ANN on a data set of 98 patients to classify the time duration of kidney condition as $<$ or $>$ 8 years to ESKD with the 85.11% average accuracy.

T.Di Noia et al. [39] designed an ensemble of ANN to estimate the risk of ESKD and the data set of 587 patients was collected during the period of 38 years. Cross-validated committees gave a maximum of 89.1% accuracy and 10-folds cross-validation with 6 neurons in each hidden layer gave the accuracy of 91.37%.

In traditional approaches, there is still a problem to learn difficult and discriminatory medical imaging features using hand-crafted and classical machine-learning facilities.

Therefore these approaches failed in complex disease applications. As can be inferred from the research works mentioned above, accuracy using clinical data is not high in general. Even in the cases where accuracy is high, the data for the same has been collected over a long period which is quite difficult. Although, deep learning (DL) has come in the form of new development, it has a large potential in various identification work in the area of image detection, speech recognition [38], computer vision [54], visual image classification [17], pluralistic image inpainting [12], medical disease detection [33], remote sensing image recognition [55], and natural language understanding [56]. Considerable success has been achieved in segmentation, classification, and detection with the help of DL in the field of pathological imaging. There are several other fields in which DL based models have already become part of the work flow with examples including dermatologist level performance for skin cancer detection [20], diabetic retinopathy [1], classification of human epithelial

-2 [23], neuro-imaging for analysis of brain tumors [40] and alzheimer disease [32], lung cancer detection [30], and breast cancer detection [49] and classification.

In bio-medical image processing, transfer learning techniques that use various pre-trained deep neural networks have also become popular and have achieved high accuracy compared to the techniques mentioned above. Some of the examples where this type of technique is used and has achieved high accuracy are mentioned below.

Ahn et al. [2] extracted X-ray image features from VGG which is a pre-trained deep convolutional neural network (DCNN) [48] and then classified X-ray with a sparse spatial pyramid model. Ahmed Alghamdi et al. [4] detected myocardial infarction (MI) based on novel deep transfer learning methods and achieved an accuracy of 99.22%, sensitivity or TPR of 99.15%, and specificity or TNR of 99.49% with VGG model. Alexander Kensert et al. [29] used ResNet50, inception−v3, and inception ResNetV2 as a pre-trained deep CNN to distinguish between different cell morphologies and obtained higher predictive accuracy in the range of 95 to 97%.

Phan et al. [41] used a pre-trained deep CNN model to extract the image features in order to classify the HEp-2 cell. Different feature selection methods are used for selecting relevant features.

IgAN's stage depends on scores of five parameters MEST-C. Predictions of all these parameters are difficult. Therefore, this research work proposes an automated system model based on a transfer learning classification model to detect the mesangial hypercellularity (M-score) in IgAN using kidney biopsy image. In this work, DL model is trained with the help of a microscopic image dataset of considerable size which contains kidney biopsy images with m_0, and m_1 score. This training model learns features from all the images and evaluates M score for IgAN. Additionally, this research work compares DL performance to different machine learning classifiers with deep features and validates MEST-C score.

The remaining work is divided into the following section: Section 2 provides the details of the image data set in which pre-processing of data is also defined and Section 3 describe the methodology in two Sections 3.1 and 3.2 respectively. Different evaluation parameters are discussed in Section 4. In Section 5, the experimental results and a discussion on the same have been presented. The future work and conclusion are drawn in Sections 6 and 7 respectively.

2 Data materials

2.1 Collection of datasets

Sonogram and x-ray images are used to locate the kidney. In kidney biopsy, small pieces of tissue are drawn with the help of a needle. These tissues are stained with periodic-acid-schiff (PAS) to highlight the required morphological features and then examined with different resolution microscopes. Eclipse E600 microscope, Nikon is used to examine the tissue under 100 × oil objective lens. DP 71 microscope digital camera is used to take images with a resolution at 1360×1024. A biopsy can help to detect kidney disease and identify the reason. For this model, kidney biopsy microscopic image dataset of different patients, which contain mesangial, endocapillary, segmental, tubular atrophy, and crescents abnormality of the glomerulus is collected. The dataset used in this work contains 138 individual glomerulus (×20 magnification digital biopsy) images of IgAN patients received from All India

Institute of Medical Science (AIIMS), Delhi. Images of normal patients have also been collected for classification from the department of pathology, AIIMS, New Delhi, India. This data is obtained after approval from the ethical committee of AIIMS.

2.2 Pre-processing of data

The process involves taking a kidney biopsy image as an input and sectioning this image into different glomerulus images (5 to 6 or more depending on the size of kidney biopsy). The sectioning process is done with the region of interest technique of image processing. The resulting images are cropped to discard undesired areas and saved in JPEG form without compression in size of 1440 $\times$ 1024 pixels of three-channel red green blue (RGB). The pathologist identifies the glomerulus and labels the region of interest (ROI) in images. Many images have been captured using 20$\times$ magnification so that the entire ROI is covered. In this model, highly professional doctors of pathology department of AIIMS annotate glomerulus with box and label the data in two classes. Before entering in classification part, some artificial methods have been used in this work to increase the number of training examples so that overfitting problem can be solved. This process is called data augmentation. Then 30% of data is used as validation from the training data.

3 Methodology

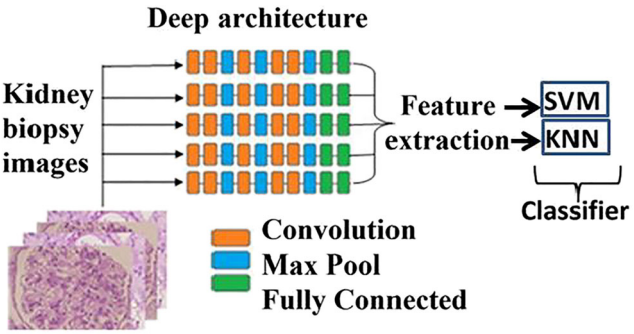
In the proposed model, the first step to identify the glomerulus is completed by the labeling process [21]. The database is divided into m_0 and m_1 score label. As shown in Fig. 3a and b, mesangial hypercellularity is identified in each glomerulus by using different machine learning (ML) classifiers with deep features or by direct transfer learning and then mesangial hypercellularity of MEST-C score is evaluated thereafter. Other abnormality in MEST-C score will be classified when sufficient data for the same is received from AIIMS.

3.1 SVM, KNN classifiers with deep features model based on mesangial hypercellularity

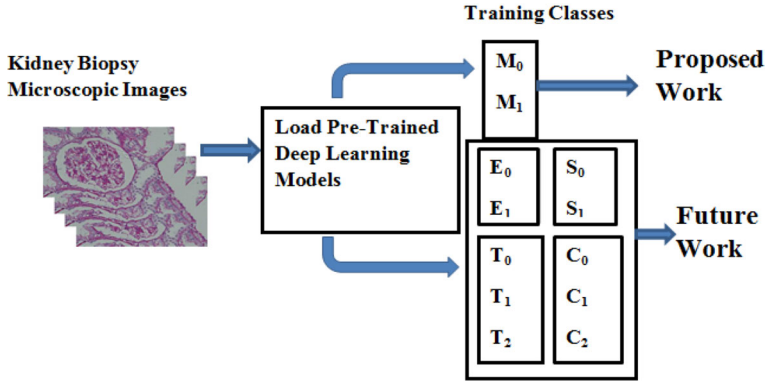
First, some basic ML models are developed to analyze the performance of the CNN models in identifying mesangial hypercellularity. Basic model architecture is shown in Fig. 3a. This work uses kidney tissue images which are given to deep architecture. This deep architecture has multiple layers and at the end, the features are extracted by a fully connected layer. These features are the input to support vector machine and k nearest neighbor classifiers associated with the respective class labels (m_0 and m_1) as output. The input data set of biopsy images is divided into training and validation by k-fold. The model is finally tested with the data set that has not been used in this process.

1. **Support vector machine (SVM):** In this technique, the training data set is divided into various classes using a boundary called hyper plane **H** [15]. However, it is not always possible to find a linear hyper plane and therefore non-linear SVM is used.
2. **k-nearest neighbors (k-NN)**

   In K-NN, the class of M score is predicted based on the minimum distance between the training data features (X_1, Y_1), (X_2, Y_2), ... (X_k, Y_k) and the new biopsy feature X_t, and that distance is called the euclidean distance (E_d).



(a) SVM, and KNN classifiers with deep feature model



(b) Deep CNN model

Fig. 3 Architecture of Mesangial hypercellularity of MEST-C score in IgAN classification

Algorithm 1 K-NN algorithms.

procedure K-NN(training data, test data)
Store the training data (X_1, Y_1), (X_2, Y_2), ... (X_j, Y_j)
Test data X_t
Find the k training examples (X_1, Y_1), (X_2, Y_2), ... (X_k, Y_k)
$i = 0$
while i≤k **do**

$$Ed = \sqrt{\sum_{n=1}^{k}((X_i - X_t)^2}$$

find kth training examples that minimize Ed
output Y_t=Y_k

3.2 DL models for detection of mesangial hypercellularity of MEST-C score in IgAN

DL models have numerous sequential sets of convolutional and sub sampling (pooling) layers [10, 46] succeeded by the supervised classifier. Each of these layers produces a set of two-dimensional feature map as an output with the help of input and weight matrix. Feature maps from one layer are used as an input to the next layer.

The architecture of this model is given in Fig. 3b. This work uses transfer learning using DL models like AlexNet, GoogLeNet, ResNet, and Inception−v3 for validation of MEST-C score of IgAN. Current work uses transfer learning because the size of mesangial patients data is small for the training of deep neural networks. Transfer learning is better as the training time of neural network models is reduced with less error. In this paper all the models are pre-trained with ImageNet dataset. This data set has 14 million images with 20 thousand categories. Transfer learning can be done by modifying last three layers of the CNN network. The initial layers of the network extract the general features of the training images where as the last FC layer learn class-specific features to classify the images into specific classes. Doctors do not have super computers to run the models, therefore transfer learning is best as it can run even in normal computers and is faster for real-time application with limited complexity of all types of hardware. In this work, input data set of the first layer in each model is the data set of biopsy images. The description of transfer learning model are given below:

1. **AlexNet Model:** This model has 25 layers which are serially linked to each other as shown in Table 1. The biopsy images are resized to 227*227 before giving them as an input to this network [31].
2. **GoogLeNet:** This model has 144 layers among which some layers are parallelly attached to each other. The biopsy images are resized to 224*224 before giving them as an input to this network [50].
3. **ResNet-50:** This model has 177 layers. This network forms a shortcut connection to skip one or more layers in the network to link to other layer [27].
4. **Inception−v3:** This model has 315 layers. Inception−v3 factorize large convolution into small convolutions to reduce the number of parameters [51].

After using multiple layers of convolutional, ReLU, and pooling layers, these models have fully connected (FC) layers which convert 2-D feature maps into a 1-D vector for classification purpose. In this model, transfer learning can be achieved by modifying last FC layer in each CNN method and replaced with two neurons FC layer because m score data set has only two output class m_0 and m_1. Classes obtained in the last classifier layer are then compared with the actual classes. Any difference is identified as an error and is sent backward for tuning the backward model parameters to reduce the error, and this process is known as back-propagation. Weights in each layer get updated by back-propagation. Training and testing processes are using biopsy-diagnose images only.

Up to the best of our knowledge of all the methods used in validation of MEST-C score of IgAN, dL is being used for the first time through this work. The data set is classified into m_0 and m_1 classes. The proposed model is described in Fig. 3b.

In the proposed method, the pre-trained models are retrained with the help of different stained images of kidney tissue. For identification of different parameters of scores like M, E, S, T, and C, the model for all the five scoring parameters is executed independently. The last fully-connected layer in the network has either 2 or 3 classes depending on number of scoring levels i.e. 0, 1 or 0, 1, 2. The same DCNN is used for all five scoring types. As discussed in the introduction section, M has two scores 0, 1; E has two scores 0, 1; S has two scores 0, 1; T has three scores 0, 1, 2; and C have three scores 0, 1, 2. However, in this study scoring is done only for mesangial hypercellularity (M) as sufficient data is not yet available for other scoring parameters. Some example images are shown in Fig. 4 for glomerulus of different patients. It is clear from the Fig. 4 that it is very difficult to identify different scoring levels for these patients, as the images share many similar visual features.

Table 1 CNN architecture

Layer	Kernel size	Stride	Pad
Input	N/A	N/A	N/A
Convolution	11	4	0
ReLU	1	1	0
Max-pooling	3	2	0
Convolution	5	1	2
ReLU	1	1	0
Max-pooling	3	2	0
Convolution	3	1	1
ReLU	1	1	0
Convolution	3	1	1
ReLU	1	1	0
Convolution	3	1	1
ReLU	1	1	0
Max-pooling	3	2	0
Fully-connected with 4096 neurons	1	1	0
Fully-connected with 4096 neurons	1	1	0
Fully-connected with 2 neurons	1	1	0

4 Evaluation parameters

Once the model is trained, it is tested using test data set and efficiency is calculated based on the test results. Test data set has to be different from the training data set in order to estimate the generalization error. Initially, the available input data set is divided into 2 data

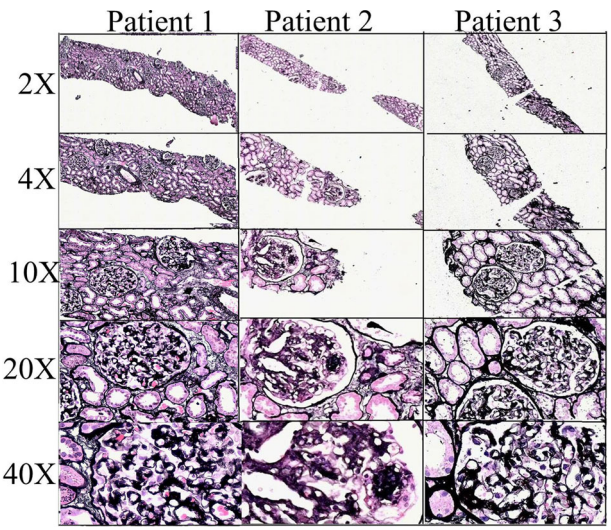


Fig. 4 Mesangial hypercellularity scoring with jone's methenamine silver stain of renal biopsy samples at different magnifications

set; one for training and the other for testing using k-fold cross-validation. The success of the model is verified using in-fold data. Testing error is calculated by taking average error of all the folds. The model can be efficiently trained with multiple fits using k-fold when small input dataset is available. Performance of all the models in this research work can be measured by confusion matrix and parameters are given below.

1.
$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}. \tag{1}$$

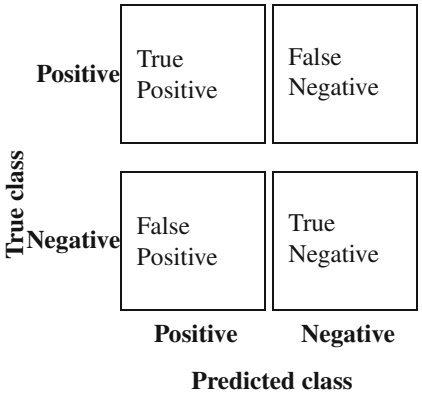
2.
$$TNR(TrueNegativeRate) = \frac{TN}{TN + FP}. \tag{2}$$

3.
$$TPR(TruepositiveRate) = \frac{TP}{TP + FN}. \tag{3}$$

4.
$$F_1score = \frac{2 * TP}{2 * TP + FP + FN}. \tag{4}$$

5. Matthews correlation coefficient (MCC)

$$\begin{aligned} MCC =& ((TP * TN) - (FP * FN))/((TP + FP) \\ & * (TP + FN) * (TN + FP) \\ & * (TN + FN))^{(0.5)}. \end{aligned} \tag{5}$$



Confusion matrix

6. Receiver operating characteristics (ROC) and area under curve (AUC): The ROC is an important graphical method to assess the performance of classifiers using false positive rate and true positive rate on X-axis and Y-axis respectively. The high value of AUC in ROC curve indicates that the model efficiently validates the mesangial hypercellularity score.

5 Results and discussion

In this research work, machine learning with deep features and transfer learning approach of CNN models are used to retrieve the information about the mesangial hypercellularity (M)

Table 2 Classification parameters of KNN, SVM, and DL models

Technique	Accuracy(%)	Sensitivity or TPR	Specificity or TNR	F_1 score	MCC
K-NN	65.3	0.45	0.85	0.56	0.33
SVM	68.7	0.39	0.98	0.55	0.46
proposed model with AlexNet	90±2	0.904	0.80	0.73	0.662
GoogLeNet	77.5	0.57	0.81	0.47	0.339
ResNet	82.4	0.71	0.841	0.52	0.45
Inception−v3	78.9	1.00	0.78	0.71	0.29

scoring parameter from the PAS-stained images of kidney biopsy using MATLAB R2020a with zero cost.

For machine learning with deep feature classification model, this work extracts features from last fully-connected layer of DL model. After feature extraction, two machine learning classification methods are used to classify the mesangial hypercellularity (M) score.

For transfer learning, this work uses Deep Network Designer application which is available in MATLAB R2020a. The process of this application consists of three parts : −

Designer, Data, and Training.

In respect of the proposed model the implementations of the above processes are described below:

1. In the designer part, this work uses four CNN models. Each model is uploaded one by one. The fully connected (FC) layer given at the end of each model is replaced with a fully connected layer with only 2 neurons due to two class label outputs. Learning rate factors are also changed in the last FC layer from 1 to 10.
2. In the data part, the biopsy data is imported. Data is further augmented using the following options:

   (a) Random reflection axis: X − axis reflection is used in the data.
   (b) Random rotation (degrees): This work uses minimum -90 and maximum of 90 degrees.
   (c) Random re-scaling: This work uses minimum 1 and maximum of 2 re-scaling.
   (d) Random horizontal translation: No change.
   (e) Random vertical translation: No change.

   Then 70 % of data is allocated for training and remaining 30 % is allocated for validation.
3. In the training process, there are many training options. This study uses stochastic gradient descent with momentum with 0.001 learning rate, and validation frequency set to 5. MaxEpochs, and MiniBatchSize is set to 10. These are the main changes done in training option according to data set. After these changes the training process is started.

For each classification task, there is also an output class which was confirmed by the doctors.

For each scoring identification, the variation in performance of the model is studied with the number of layers, the number of epoch, batch size and many of such sensitive parameters. The DCNN model which gives the best results is chosen.

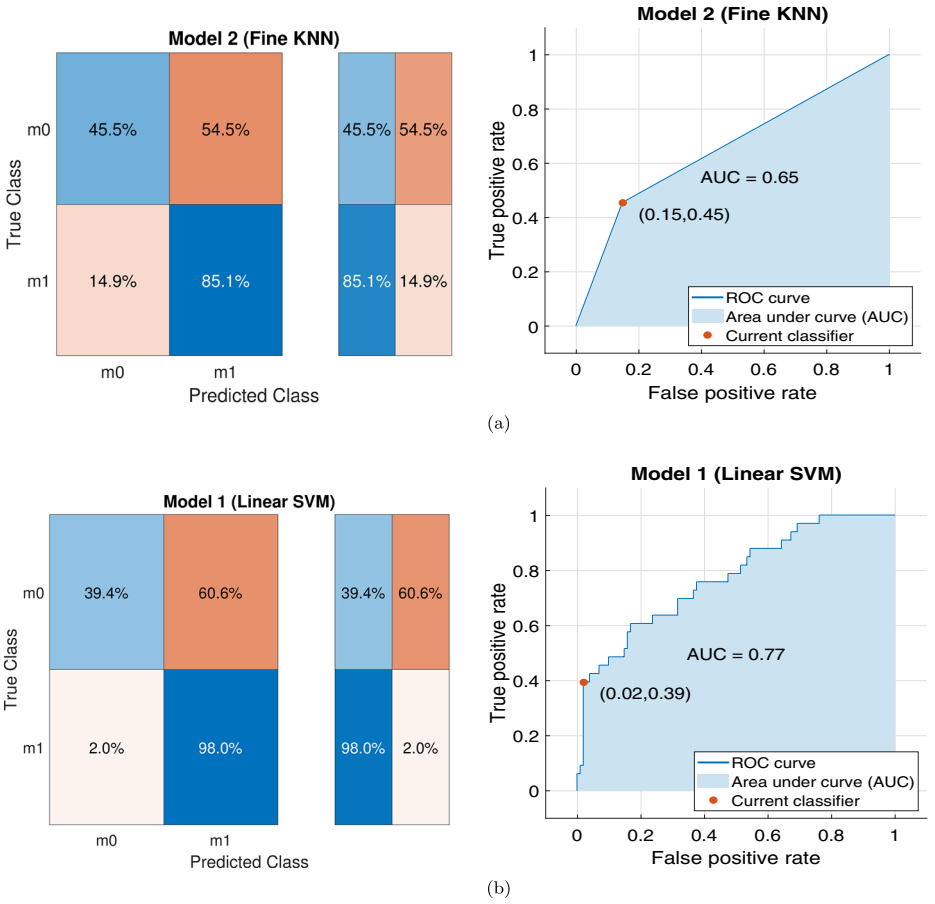


Fig. 5 Confusion matrix & ROC curve by KNN and SVM classifiers trained with deep features

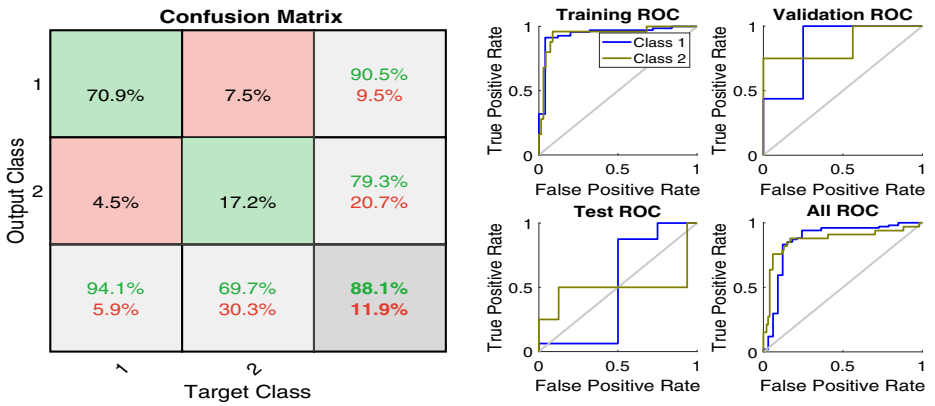


Fig. 6 Confusion matrix and ROC curve of AlexNet DL model

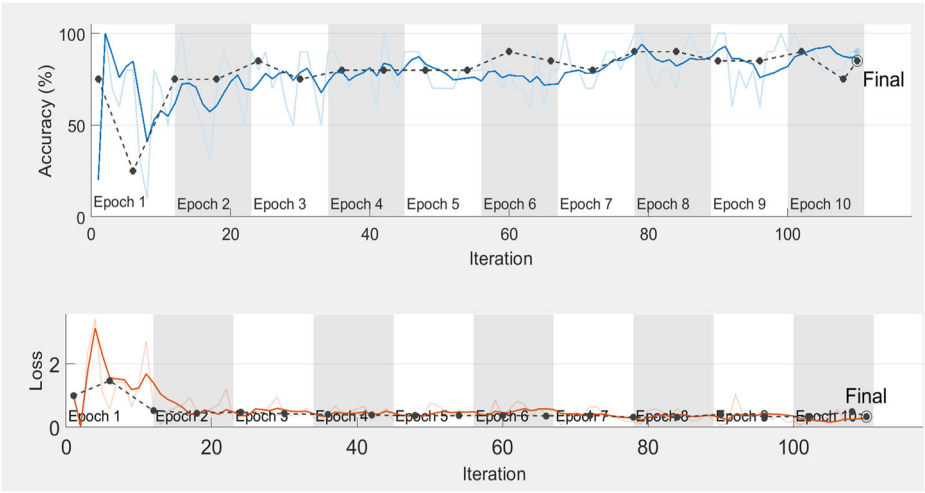


Fig. 7 AlexNet model performance

As a result of the CNN architecture, even single magnification of images performed better for the model. In this work, a deep model is trained using one set of $\times 20$ magnification images and tested using another set of $\times 20$ magnification images.

The success of the model can be checked from the accuracy, ROC curve, and loss curve. The loss curve tends to be stable after descending, indicating that the training processes converge. In this study, all the patients who have been diagnose with IgA after kidney biopsy are included.

Other metrics like accuracy, sensitivity or TPR , specification, F1 score, and MCC have been calculated to confirm the superiority of the DCNN model of mesangial hypercellularity over other models.

Figure 5 shows the KNN and SVM performance parameters. Figures 6 and 7, shows AlexNet performance parameters when AlexNet model applied as a pre-trained network in Fig. 3b. In Fig. 3b, AlexNet is replaced with GoogLeNet to get results in Figs. 8 and 9; is

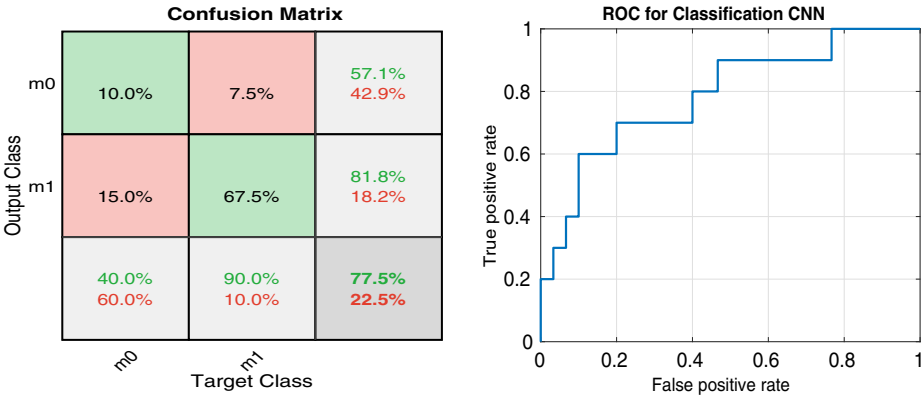


Fig. 8 Confusion matrix and ROC curve of GoogLeNet DL model

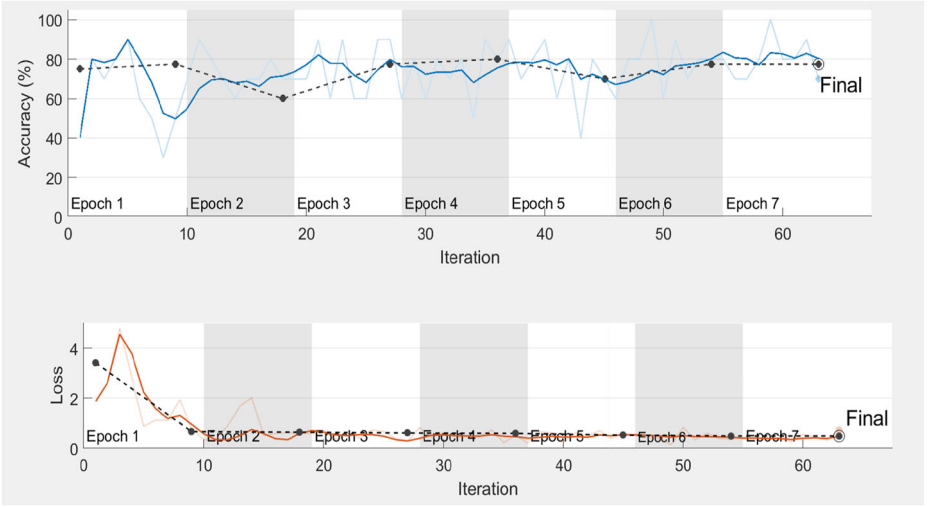


Fig. 9 GoogLeNet model performance

replaced with ResNet to get results in Figs. 10 and 11; and is replaced with Inception−v3 to get results in Figs. 12 and 13. The proposed model is different from the other models in terms of accuracy and loss curve when it comes to small data set and time taken in training the model. Table 2 shows the comparison of all the techniques which were applied to predict the mesangial hypercellularity score of IgA. It can be seen from Table 2 that the highest accuracy is obtained with the AlexNet model. The accuracy up to 89±2% in identifying the mesangial hypercellularity score of IgA is obtained using AlexNet architecture. Other CNN models also give the respectable performance but the AlexNet model is superior over other classification models. When the loss figures of all the models are analyzed, it is clear that AlexNet and GoogLeNet model decreases the value faster and approaches zero, but GoogLeNet has less accuracy than AlexNet. So, it is clear that the AlexNet model is superior over the other classification models in terms of accuracy and loss.

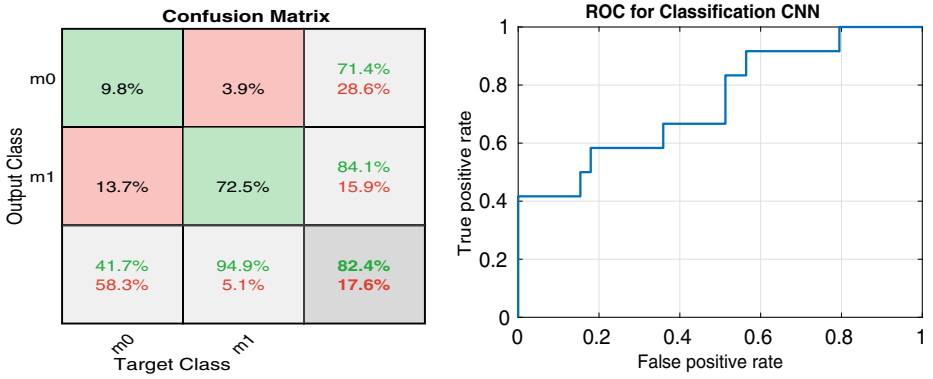


Fig. 10 Confusion matrix and ROC curve of ResNet DL

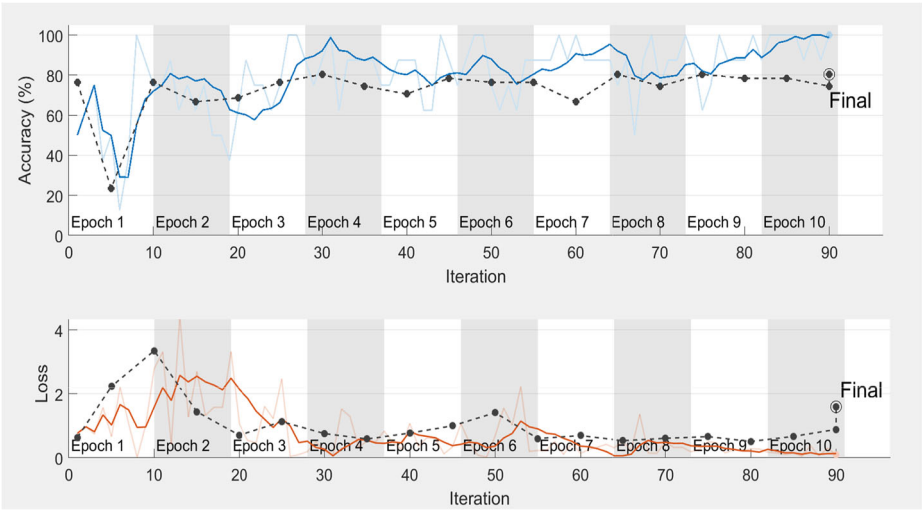


Fig. 11 ResNet model performance

As can be seen from Table 3 which is the summary of the literary works mentioned in the introduction section, accuracy using clinical data has not been high in general. Even in cases where accuracy is high, it takes a long period to collect the clinical data which is quite impractical. In proposed model data is small and it took 1.5 years for collection, which is quite feasible.

Recently DCNN is being used in medical field for image based classification of clinical subjects. The problem of hematology [42], radiology, dermatology [20], oncology, and many such clinical topics are solved by the DCNN model.

As of now, no other work similar to intensive learning approach in IgAN has been published.

It is assumed that the proposed model will predict other parameters of MEST-C scoring namely endocapillary hypercellularity, segmental Sclerosis, tubular atrophy, and crescent in

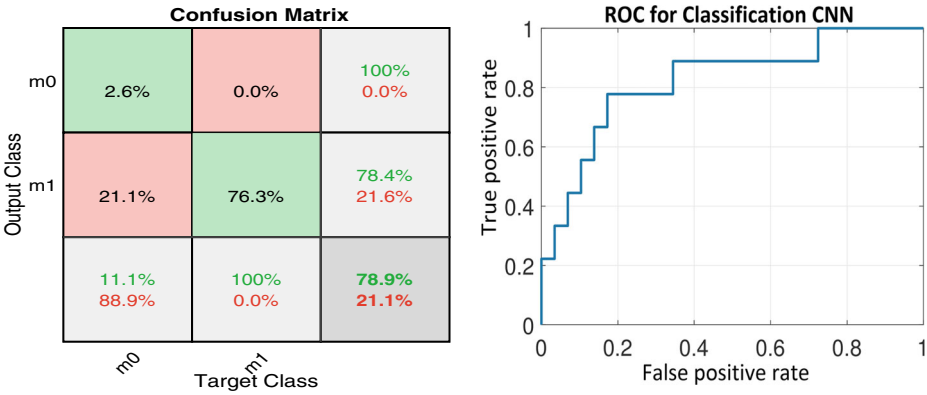


Fig. 12 Confusion matrix and ROC curve of Inception−v3 DL

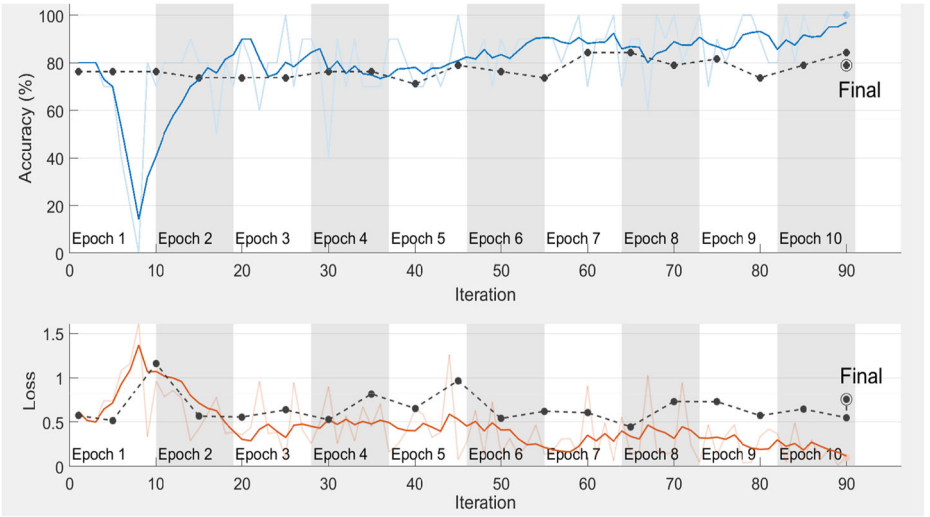


Fig. 13 Inception−v3 model performance

the same manner as above. The model will be trained for these parameters once sufficient data is available for the same.

5.1 Advantages

This model has many advantages. This model uses transfer learning and the process of transfer learning has promoted the use of DCNN models for classification function with small data sets. Transfer learning has removed the correlation between the large dataset size and high model accuracy. This because transfer learning uses CNN model which is already trained with millions of non-medical images. Apart from this, transfer learning also takes less training time because the pre-trained CNN does not require tuning of all the parameters.

Table 3 A comparison of different studies in IgAN based on number of patients, collection year and performance

Author (Year)	Method	Number of patients	Collection time	Performance
Geddes et al. [24]	artificial neural network	54	10Yr	Accuracy: 87.0%, Sensitivity or TPR : 86.4%, Specificity or TNR : 87.5%
Cannone et al. [13]	naive Bayes, SVM, decision tree, and ANN	98	NA	Accuracy: 85.11%
T.Di Noia et al. [39]	ensemble of ANN	587	38Yr	Accuracy 89.1%, 91.37%.
Prposed model	CNN	20	1.5Yr	Accuracy 90±2%, Sensitivity or TPR : 90%, Specificity or TNR : 80%

One of the strengths of this work is that it displays the potential of ML algorithm in detecting pathological diseases using the diagnostic images of the patients. The results obtained by running the algorithm are in accord with clinical results. The accuracy of the algorithm is at least equivalent to that of an experienced nephrologist, who is able to achieved the same only after a labour intensive morphological investigation.

If this model is developed into an easily accessible mobile or a computer application, then the results can be immediately obtained by providing histological images as an input. The work flow for above application will be as follows:

1. Pathology clinic conducts kidney biopsy on the suspecting patients and generates an RGB color image of the biopsy samples containing at least 5 to 6 glomerulus.
2. Images of individual glomerulus are extracted from the biopsy images either manually or using ROI.
3. Each glomerulus image is then fed into the application of the trained CNN model installed on the doctor's system. The model will immediately detect whether mesangial hypercellularity is present in that glomerulus.
4. Based on the results obtained for all the glomerulus present in one patients biopsy image, M scoring of MEST-C score can be done.
5. The M score obtained will help in determining the stage and severity of IgAN so that appropriate care and treatment can start right away.

The above work flow is especially useful in the areas where nephrology experts are not easily available. In these areas, even general pathologist can immediately obtain results using this model and can prepare a clinical report accordingly.

6 Future work

1. Future models similar to the proposed model for calculating M score will be developed for calculating other parameters of MEST-C score namely endocapillary hypercellularity , segmental sclerosis, tubular atrophy, and crescents.
2. For all pathology cases, the model will be further developed into a diagnostic tool, which will predict results immediately using kidney biopsy images of a patient without the need of any human diagnoses.
3. The dataset used in my proposed work is different from general data-sets because it belongs to a particular disease. It consists of images of multiple glomerulus with the cluttered background of kidney tissue, which reflects its complexity. As paper [21] describes that containing multiple salient objects in an image creates a lots of clutter. That's why in the proposed model, high-quality labelling is used which plays an important role to improve accuracy. In this model highly professional doctor of pathology annotate the glomerulus with box. In the proposed model first step of identifying the glomerulus will be completed by the labelling process that is described above and as in paper [21].
4. I will try to incorporate Deep Side idea in my research work in future. I will design a Deep Side structure with different depth in AlexNet and concatenate the side networks by using Deep Side nonlinear process so that it provides boundary details of the tissue structure. As mention in paper [22] that Deep Side nonlinear achieves better result and also learns faster than other networks, so it will also be a good choice for my work.
5. Local edge information of Mesangial area also plays a very important role for better prediction of M score. As shown in paper [57], Edge Guidance Network is used for

the detection of edge features. In future, this work can be implemented in the proposed model to obtain better predictions.

The above work will be implemented through the following steps:

(a) Identify glomerulus by using labelling process that is described in point 3 above.
(b) Use basic CNN model and connect its deep side path for boundary and hierarchical features.
(c) The initial convolution layer of the CNN network preserves better edge features. Fusion of these edge features with boundary and hierarchical features can achieve better prediction results.

7 Conclusion

The stages of IgAN depends on scores of five parameters MEST-C. The predictions of all these parameters are difficult. An automated system model based on a transfer learning to detect the mesangial hypercellularity (M-score) in IgAN using kidney biopsy images has been developed in this research work. The pre-trained DL models (AlexNet, GoogLeNet, ResNet, and Inception−v3) were used to learn complex features and fine-tuned for biopsy data. The proposed model has been compared with the basic classifiers which are trained with deep features. The proposed model outperforms the classification performance than other classifiers. AlexNet transfer learning model used in this analysis gives the highest performance parameter with 90±2% accuracy, 90.4% sensitivity and 80% specificity. A high accuracy with small data set and little collection time has been obtained by transfer learning.

Acknowledgements The authors are grateful to department of pathology, All India Institute of Medical Sciences, New Delhi, India, for sharing the dataset. We express gratitude towards Professor of pathology department Dr. A.K. Dinda and his team for the data used in this research and ethical permission number is IEC-48/02.02.2018.

References

1. Abbas Q, Fondon I, Sarmiento A, Jiménez S, Alemany P (2017) Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. Med Biol Eng Comput 55(11):1959–1974
2. Ahn E, Kumar A, Kim J, Li C, Feng D, Fulham M (2016) X-ray image classification using domain transferred convolutional neural networks and local sparse spatial pyramid. In: IEEE 13th ISBI, pp 855–858
3. Alamartine E, Sabatier JC, Guerin C, Berliet JM, Berthoux F (1991) Prognostic factors in mesangial IgA glomerulonephritis: an extensive study with univariate and multivariate analyses. Am J Kidney Dis 18(1):12–19
4. Alghamdi A, Hammad M, Ugail H, Abdel-Raheem A, Muhammad K, Khalifa HS, Abd El-Latif AA (2019) Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. Multimed Tools Appl 1–22
5. Alghamdi AS, Polat K, Alghoson A, Alshdadi AA, Abd El-Latif A (2020) A novel blood pressure estimation method based on the classification of oscillometric waveforms using machine-learning methods. Appl Acoust 164
6. Alghamdi AS, Polat K, Alghoson A, Alshdadi AA, Abd El-Latif A (2020) Gaussian process regression (GPR) based non-invasive continuous blood pressure prediction method from cuff oscillometric signals. Appl Acoust 164

7. Avci E (2009) A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier. Expert Syst Appl 36(7):10618–10626
8. Barbour SJ, Espino-Hernandez G, Reich HN, Coppo R, Roberts IS, Feehally J, Herzenberg AM, Cattran DC, Bavbek N, Cook T, Troyanov S (2016) The MEST score provides earlier risk prediction in lgA nephropathy. Kidney Int 89(1):167–175
9. Bartosik LP, Lajoie G, Sugar L, Cattran DC (2001) Predicting progression in IgA nephropathy. Am J Kidney Dis 38(4):728–735
10. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828
11. Berthoux F, Mohey H, Laurent B, Mariat C, Afiani A, Thibaudin L (2011) Predicting the risk for dialysis or death in IgA Nephropathy. J Am Soc Nephrol 22(4):752–761
12. Cai W, Wei Z (2020) PiiGAN: generative adversarial networks for pluralistic image inpainting. IEEE Access 8:48451–48463
13. Cannone R, Castiello C, Fanelli AM, Mencar C (2011) Assessment of semantic cointension of fuzzy rule-based classifiers in a medical context. In: 11th international conference on intelligent systems design and applications, Cordoba, pp 1353–1358
14. Cattran DC, Coppo R, Cook HT, Feehally J, Roberts IS, Troyanov S, Alpers CE, Amore A, Barratt J (2009) The Oxford classification of IgA nephropathy: rationale, clinicopathological correlations, and classification. Kidney Int 76(5):534–545
15. Cristianini N, Shave-Taylor J (2000) An introduction to support vector machine and other kernel-based learning methods. Cambridge University Press, New York
16. Cruz-Ramírez M, Hervás-Martínez C, Fernández JC, Briceno J, delaMata M (2013) Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks. Artif Intell Med 58(1):37–49
17. El-Rahiem BA, Ahmed MAO, Reyad O, El-Rahaman HA, Amin M, El-Samie FA (2019) An efficient deep convolutional neural network for visual image classification. In: International conference on advanced machine learning technologies and applications, pp 23–31
18. Escudero J, Ifeachor E, Zajicek JP, Green C, Shearer J, Pearson S (2013) Machine Learning-Based Method for Personalized and Cost-Effective Detection of Alzheimer's Disease. IEEE Trans Biomed Eng 60(1):164–168
19. Esposito M, DeFalco I, DePietro G (2011) An evolutionary, fuzzy DSS for assessing health status in multiple sclerosis disease. Int J Med Inform 80(12):245–254
20. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM (2017) Sebastian Thrun Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118
21. Fan DP, Liu J, Gao S, Hou Q, Borji A, Cheng MM (2018) Salient objects in clutter: bringing salient object detection to the foreground. CoRR
22. Fu K, Zhao Q, Gu IreneYu-Hua, Yang J (2019) Deepside: A general deep framework for salient object detection. Neurocomputing 356:69–82
23. Gao Z, Zhang J, Zhou L, Wang L (2014) HEp-2 cell image classification with CNN. In: Pattern recognition techniques for indirect immunofluorescence images (I3A). 1st IEEE Workshop on 24–28
24. Geddes C, Fox J, Allison M, Boulton-Jones J, Simpson K (1998) An artificial neural network can select patients at high risk of developing progressive IgA nephropathy more accurately than experienced nephrologists. Nephrol Dial Transplant 13(1):67–71
25. Goto M, Wakai K, Kawamura T, Ando M, Endoh M, Tomino Y (2009) A scoring system to predict renal outcome in IgA nephropathy: a nationwide 10-year prospective cohort study. Nephrol Dial Transplant 24(10):3068–3074
26. Gray H (1918) Anatomy of the human body, vol 8. Lea & Febiger, Philadelphia
27. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. CoRR
28. Kassahun Y, Perrone R, DeMomi E, Berghöfer E, Tassi L, Canevini MP, Spreafico R, Ferrigno G, Kirchner F (2014) Automatic classification of epilepsy types using ontology-based and genetics-based machine learning. Artif Intell Med 61(2):79–88
29. Kensert A, Philip JH, Ola S (2019) Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. SLAS 24(4):466–475
30. Kirienko Ma, Sollini M, Silvestri G et al (2018) Convolutional neural networks promising in lung cancer t-parameter assessment on baseline FDG-PET/CT. Contrast Media Mol Imaging
31. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: NIPS'12 proceedings of the 25th international conference on neural information processing systems, pp 1097–1105
32. Liu M, Cheng D, Yan W (2018) Classification of alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. Front Neuroinform 12–35

33. Lundervold A, Lundervold A (2019) An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik 29(2):102–127
34. Ma Y, Chen W, Ma X, Xu J, Huang X, Maciejewski R, Anthony KHT, EasySVM (2017) A visual analysis approach for open-box support vector machines. Comput Vis Med 3(2):161–175
35. MacKinnon B, Fraser EP, Cattran D, Fox JG, Geddes CC (2008) Validation of the Toronto formula to predict progression in IgA nephropathy. Nephron Clin Pract 109(3):148–153
36. Maglogiannis E, Anagnostopoulos I (2009) Zafiropoulos An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. Appl Intell 30(1):24–36
37. Mandal I, Sairam N (2013) Accurate telemonitoring on Parkinson's disease diagnosis using robust inference system. Int J Med Inform 82(5):359–377
38. Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K (2019) Speech recognition using deep neural networks: a systematic review in IEEE access 7:19143–19165
39. Noia TD, Ostuni VC, Pesce F, Binetti G, Naso D, Schena FP, Sciascio ED (2013) An end stage kidney disease predictor based on an artificial neural networks ensemble. Expert Syst Appl 40(11):4438–4445
40. Pereira S, Pinto A, Alves V, Silva CA (2016) Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging 35(5):1240–1251
41. Phan HTH, Kumar A, Kim J, Feng D (2016) Transfer learning of a convolutional neural network for HEp-2 cell image classification. In: IEEE 13th ISBI, pp 1208–1211
42. Purwar S, Tripathi RK, Ranjan R, Saxena R (2019) Detection of microcytic hypochromia using cbc and blood film features extracted from convolution neural network by different classifiers. Multimed Tools Appl 79:1–23
43. Radford MG, Donadio JV, Bergstralh EJ, Grande JP (1997) Predicting renal out-come in IgA nephropathy. J Am Soc Nephrol 8(2):199–207
44. Rauta V, Finne P, Fagerudd J, Rosenlof K, Tornroth T, Gronhagen Riska C (2002) Factors associated with progression of IgA nephropathy are related to renal function-a model for estimating risk of progression in mild disease. Clin Nephrol 58(2):85–94
45. Roychowdhury S, Koozekanani DD, Parhi KK (2014) DREAM: diabetic retinopathy analysis using machine learning. IEEE J Biomed Health Inform 18(5):1717–1728
46. Scherer D, Müller A, Behnke S (2010) Evaluation of pooling operations in convolutional architectures for object recognition. In: 20th International Conference on Artificial Neural Networks (ICANN), pp 92–101
47. Sheppard D, McPhee D, Darke C, Shrethra B, Moore R, Jurewitz A, Gray A (1999) Predicting cytomegalovirus disease after renal transplantation : an artificial neural network approach. Int J Med Inform 54(1):55–76
48. Simonyan K (2014) Very deep convolutional networks for large-scale image recognition. Comput Vis Pattern Recognit
49. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2016) Breast cancer histopathological image classification using convolutional neural networks. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp 2560–2567
50. Szegedy C, et al. (2015) Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp 1–9
51. Szegedy C, Vanhoucke V, Ioffe S, Shlens J (2016) Rethinking the inception architecture for computer vision
52. Tenório JM, Hummel AD, Cohrs FM, Sdepanian VL, Pisa IT, de Fátima Marin H (2011) Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease. Int J Med Inform 80(11):793–802
53. Tortora GJ, Derrickson B (2017) Principles of anatomy and physiology. Wiley, New York
54. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review in computational intelligence and neuroscience
55. You H, Tian S, Yu L, Lv Y (2020) Pixel-level remote sensing image recognition based on bidirectional word vectors. IEEE Trans Geosci Remote Sens 58(2):1281–1293
56. Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing [review article]. IEEE Comput Intell Mag 13(3):55–75
57. Zhao JX, Liu J, Fan DP, Cao Y, Yang J (2019) EGNet: edge guidance network for salient object detection. In: IEEE international conference on computer vision

Affiliations

Shikha Purwar[1] · Rajiv Tripathi[1] · Adarsh Wamanrao Barwad[2] · A. K. Dinda[2]

[1] Electronics and Communication Department, National Institute of Technology Delhi, Delhi, India
[2] Pathology Department, All India Institute of Medical Sciences, New Delhi, India