Predictor Model for Chronic Kidney Disease using Adaptive Gradient Clipping with Deep Neural Nets

Neeraj Sharma, Praveen Lalwani
School of Computing Science and Engineering
VIT Bhopal University

***Abstract*—This research aims to develop computer vision based predictive model for the three prominent kidney ailments namely Cyst, Stone, and Tumor which are common renal disorders that require timely medical intervention. This classification model is tested and trained using the multi-class CT Kidney Dataset which contains 12,446 images collected from PACS (Picture Archiving and Communication System) from different hospitals in Dhaka, Bangladesh. Initial models are build using plain VGG16, ResNet50, and InceptionV3 deep neural nets. Then after clip value filter of ADAM optimizer is applied which results in marginally improved accuracy and at the last Adaptive Gradient Clipping is applied as a replacement of batch norm process and this produces overall best results. The Adaptive Gradient Clipping based model achieves accuracy of 97.15% in VGG16, 99.5% in ResNet50, and 99.23% in InceptionV3. Overall classification metrics are best for ResNet50 and Inception V3 with Adaptive Gradient Clipping technique.**

***Keywords*—CT Kidney; VGG16; ResNet50; InceptionV3; gradient clipping; image processing; multiclass classification**

I. Introduction

Computer vision is emerging as one of the most promising solution for early diagnosis and assisting doctors, medical health professionals to reduce the work load and provide treatment on need basis instead of first come first serve basis. Deep learning models for medical image classification has gained lot of popularity recently because of highly accurate results generated by these models. Post COVID-19 the availability of large image data sets has also gained popularity and this promotes the deployment of deep neural nets for classification purpose. But as discussed in [1] [1] the traditional way of handling image datasets may not work well for handling the large image datasets. This research work uses CT Kidney image dataset [2], which is a collection of 12,446 images spread across four categories of Normal, Cyst, Stone, and tumor, and build convolutional neural net based multi-class classification models for CT Kidney dataset using VGG16, ResNet50, and InceptionV3 deep neural nets. The initial models are build using plain versions of VGG16, ResNet50, and InceptionV3. To build the second version the clip norm filter of Adam Optimizer is used which is a floating value and is basically used to individually clip the gradient of each weight so that the norm of each weight remains less than or equal to this value. To build the third version the Adaptive Gradient Clipping techniques as discussed in [3] is used. Adaptive Gradient Clipping is a replacement to batch norm technique when we want to train models using larger batch sizes and increase the learning rate of model.

A. Chronic Kidney Disease

Chronic Kidney Disease (CKD) is a condition in which the kidneys are impaired and cannot filter blood as effectively as they could. As a result, the body stores excess fluid and blood waste, which can contribute to a variety of health problems such as heart disease and stroke. CKD is also reported to lead to Kidney Failure and is estimated to be present in 1 out of 10 adults [4]. CKD effects almost 1 billion people worldwide [5] largely including women, older citizens, and people suffering from diabetes and hypertension. CKD causes premature morbidity and mortality and lowers quality of life; it is also expensive [6] and becomes a big financial burden for low and middle income countries.

Cysts, Tumor, and Stone are three different impairments in Kidney. Round fluid-filled pouches called kidney cysts can develop on or inside the kidneys, impairing their regular function. A growth or collection of abnormal cells that develops on the kidney is called a kidney tumor. Malignant (cancerous) or benign (not cancerous) terms might apply to these tumors. Hard deposits of minerals and salts that accumulate inside the kidneys are called kidney stones.

CKD meets all the four criteria that are required to be recognize a disease as a public health issue [7] [8]. Early detection of CKD may prevent death and disability but such early detection is difficult [9] [10] depends on the availability of nephrologists who are scarcely available in various geographic locations and specially in South Asia[11]. Delay in detection of Kidney cysts, stones, and tumors increases the possibility of renal failure [12]. All these condition pave way for deployment of Deep Neural Net based models for timely detection of Kidney related diseases.

B. Authors Contribution

Chronic Kidney Disease is recognised as a public health issue but remains a less explored disease in the medical image processing filed. In this article a reliable and automated Kidney disease image processing model using advanced deep learning models is build. Another contribution is to explore the two different approaches for image classification: the traditional way using data scaling and Batch Normalization, and the novel way of using adaptive gradient clipping. The third contribution is in comparing the performance of the three most contemporary deep learning models namely VGG16, ResNet50, and Inception V3.

C. Article Organization

In Section 2, the most prominent and related work on deep learning based image processing models are discussed and the most prominent work in this field is summarized. The preliminaries of the image processing field are discussed in Section 3. In Section 4, the dataset, preprocessing methods, and the proposed model is discussed. In Section 5 the experimental results are discussed and comparative study of the different approaches are done. The article concludes by summarizing this research work and stating the future research prospects in Section 6.

II. Related Work

A nice survey on medical image analysis can be found in [13]. Deep Learning methods have been tested on plethora of data sets in many research articles in the last decade. Main focus area in these research articles were radiology findings and segmentation tasks. Convolutional Neural Networks (CNN) have dominated the image processing segment [14], [15]. ResNet [14] has proved very helpful in building efficient Deep Neural Net Models. Other Deep Neural Net models prominently used are inception[15], and exception [16]. Another recent Deep Learning model is EfficientNet [17].

Recent advancement in the Deep Learning field like transfer learning is also proving very helpful in developing efficient models. Transfer learning is profoundly used for Natural Language Processing. Weights of a deep neural network pretrained on a large generic dataset are used to initialize subsequent tasks which can be solved with fewer data points, and less compute [18] [19] [20]. Transformers were proposed in [21] for machine translation and much recently they are deployed in image processing applications also. Multiple works try combining CNN-like architectures with self-attention [22] [23] [24], some replacing the convolutions entirely [25] [26]. But in large-scale image recognition, classic ResNetlike architectures are still state of the art [27] [28] [29][30].

Models like Vision transformer [31] have produced wonderful results as compared to Convolutional Neural Networks but hybrid models [32] and CNN with novel optimization techniques are expected to achieve much better results [33]. Deep neural nets use Gradient Descent as the basic technique for learning and a comparison of the different implementations of Gradient Descent is discussed in [34]. In all the models ADAM optimizer is used which is combination of RMSprop and Stochastic Gradient Descent with momentum. In [35] Stochastic Gradient Descent (SGD) is stated to perform better than ADAM, and RMS Prop optimizer using ResNet50. Class Balancing is discussed as a future work in this article but [36] discusses the meta-heuristic approaches that can be used for balancing X-Ray image datasets.

Limitation of Batch Norm technique are discussed in [37] for training large datasets and also discussed are the solution in the form of an novel Robust Normalization technique which provides all benefits of Batch Norm while mitigating the adversarial attacks. The technique of Adaptive Gradient Clipping is introduced in [3] as a replacement of Batch norm. Performance of Swin Transformers is reported to excel CNN models in [2] as it achieves accuracy of 99.30%. Vision Transformers are compared with CNN in [31] and it is concluded then though CNN are slower as compared to Vision Transformers because of pooling operation but Vision Transformers require large data sets for training. YOLOv8 is deployed on CT Kidney Dataset for multi-class classification in [38] and it achieves an accuracy of 82.52

III. Preliminaries

The underlying concepts of Image Processing, various classification techniques, and optimization techniques involved in this research article are discussed in this section.

A. Medical Image Processing

Medical image processing is used by medical practitioners to detect and diagnose diseases at early stages to increase chances of curing the disease. Out of the several basic stages of image processing Deep Learning is prominently applied for Image Enhancement, Segmentation and classification stages.

Noise frequently deteriorates medical pictures because of a variety of interference sources and this interfere with image processing systems' measuring procedures raising the requirement of Image Enhancement procedures [39]. The technique of segmenting a picture involves breaking it up into areas with similar texture, color, brightness, contrast, and gray level [40]. A segmented image become more meaning full and facilitates classification. Image classification is the task of categorizing an image into either of the given categories. Deep learning excels in the tasks of Image Segmentation, Reconstruction, and Classification [41]. Deep Learning models like RESNet [42], VGG16 [43], Inception Net [44] have excelled in this field achieving results that even surpass human abilities.

B. Image Classification Models

VGG16, ResNet50, and InceptionNetV3 are used for building the multi-class classification model.

1) VGG16: VGG16 [43] has become popular because of its proficiency in image classification tasks. It is a 16 layer Convolutional Neural Network (CNN) which has shown impressive performance on various image classification benchmarks. Its structure includes multiple convolutional and pooling layers followed by fully connected layers as shown in figure. It has a deep architecture to learn hierarchical features and the use of small 3x3 convolutional filters enables capturing fine-grained details in images. VGG16 has been employed in numerous studies for diabetic retinopathy and kidney cyst classification tasks, achieving high accuracy and demonstrating its potential as a reliable model in computer vision-based prediction. Structure of VGG16 classifier model is shown in Fig. 1.

2) ResNet50: ResNet50, short for Residual Network with 50 layers, is variant of a powerful CNN architecture introduced in [42] that addresses the challenge of training very deep neural networks. It introduces skip connections, or residual connections, which enable the network to learn residual mappings, making it easier to optimize and alleviate the vanishing gradient problem. ResNet50 has demonstrated superior performance on various image classification tasks, including diabetic retinopathy and kidney cyst classification. It's deep architecture allows for capturing intricate features, leading to

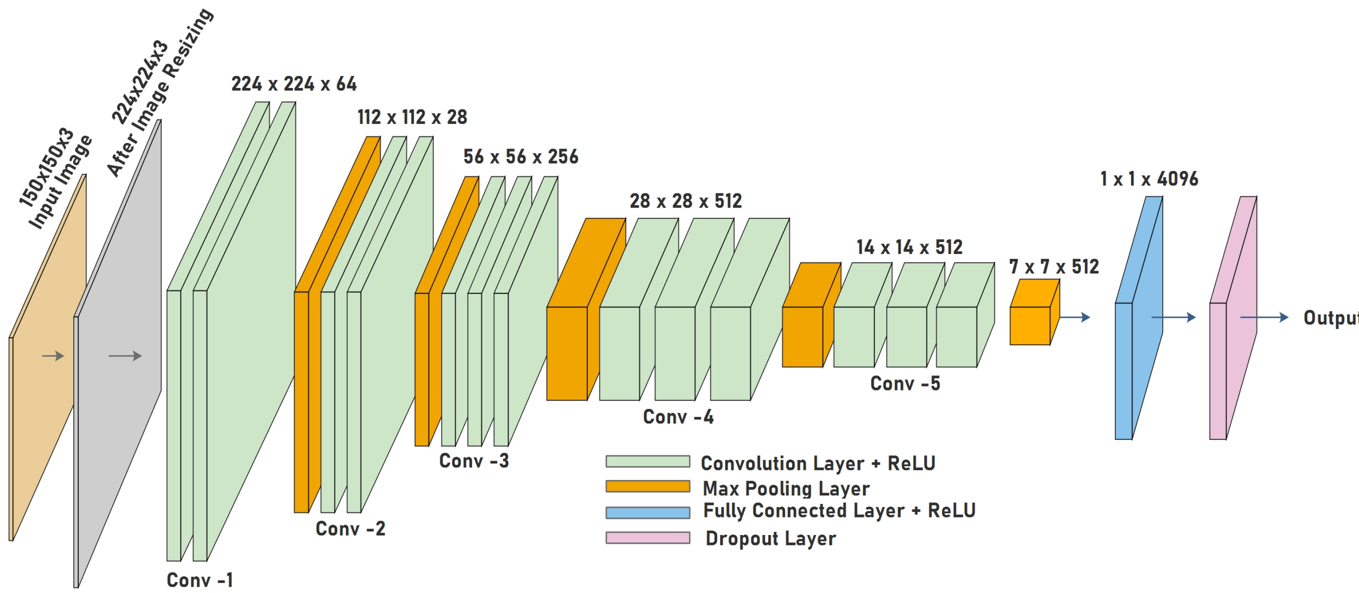

Fig. 1. Structure of VGG16.

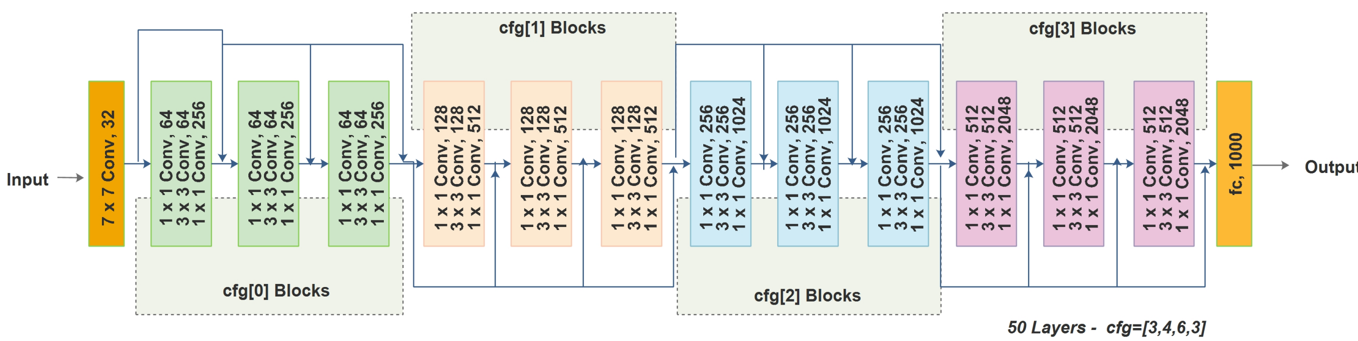

Fig. 2. Structure of ResNet50.

accurate predictions. The inclusion of residual connections makes ResNet50 particularly effective in handling complex visual patterns and has been widely adopted in computer vision research. The architecture of ResNet50 is shown in Fig. 2. cfg[3,4,6,3] refers to (3x3=9), (3x4=12), (3x6=18), and (3x3=9) summing to 48 layers (9+12+18+9) with one input layer and one fully connected layer increasing the count to 50. Structure of ResNet50 classifier is shown in Fig. 2.

3) InceptionV3: InceptionV3 is an advanced CNN architecture that incorporates the concept of "Inception modules" introduced in [44] These modules use different filter sizes and perform parallel convolutions, using which the features are captured by network at multiple scales. InceptionV3 strikes a balance between depth and computational efficiency, achieving high accuracy while maintaining a manageable model size. It has been successfully applied in various image classification tasks, including diabetic retinopathy and kidney cyst classification. InceptionV3's ability to capture both global and local features, along with its efficient architecture, makes it a valuable tool in computer vision-based prediction tasks. Structure of InceptionV3 classifier is shown in Fig. 3.

C. Normalization Techniques

To enhance the contrast of image and to make image features more visible the contrast is increased by performing min max scaling of images which scales the pixel values to a specific range.

Batch normalization [45] is frequently used for training deep neural networks with several benefits as discussed in [46] [47] [48] [49] [50]. But batch normalization comes with a major drawback of higher memory and time overheads and a mismatched behavior of training model and interference model [51] [52]. Another important limitation of Batch Normalization is that they break the interdependence between training examples. Other limitations of batch normalization are discussed in [53] [54] [55] [56] [57].As a result Normalization Free ResNets (NFRN) were developed and discussed in [46] [47] [58] and these NFRN were made more efficient by using additional regularization mechanisms as discussed in [46] [47].

Gradient Clipping: Gradient Clipping technique was introduced in [59], and the related benefits are demonstrated in [60] which uses gradient clipping to stabilize training in

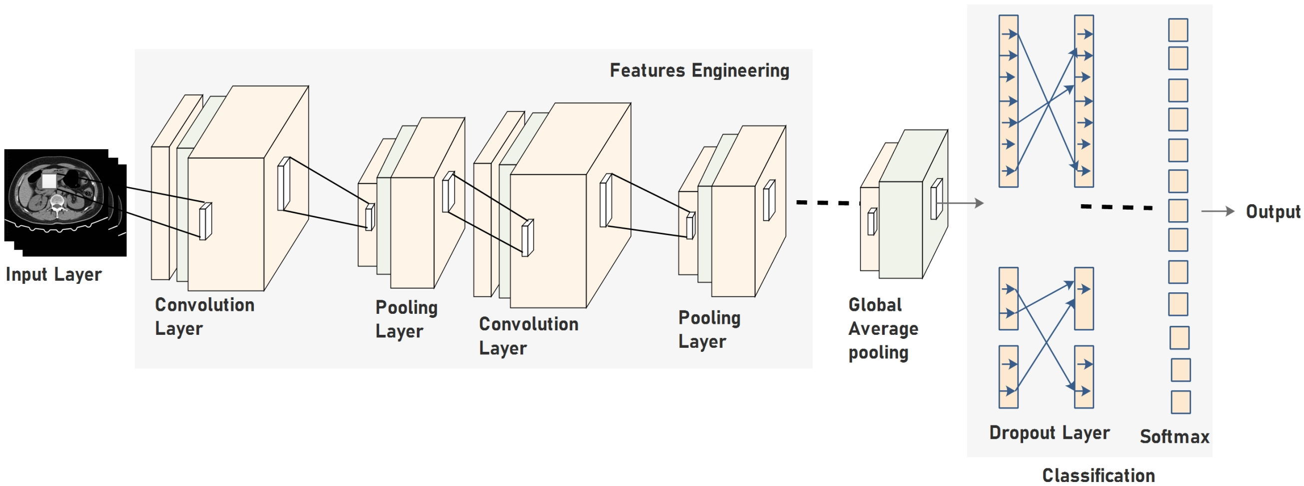

Fig. 3. Structure of InceptionV3.

Algorithm 1 Adaptive Gradient Clipping

Input:
Model parameters: θ
Rate of learning: α
Clipping threshold: ϵ
Scaling factor: γ
Initialization:
Set the model's initial values(parameter) θ
Set the initial scaling factor γ (it is usually set to 1.0)
Repeat for every training iteration::
    Compute gradients of the loss function concerning model parameters: ∇_θloss
    Compute the norm of the gradients: $||\nabla_\theta \text{loss}||$
if $||\nabla_\theta \text{loss}|| > \epsilon$ **then**
    Update the scaling factor γ: $\gamma = \frac{\epsilon}{||\nabla_\theta \text{loss}||}$
end if
Clip gradients: $\nabla_\theta \text{loss} = \gamma \cdot \nabla_\theta \text{loss}$
Update model parameters using the clipped gradients: $\theta = \theta - \alpha \cdot \nabla_\theta \text{loss}$

Large Language Models. Gradient clipping is mainly used to counter the Exploding Gradient and Vanishing gradient problems. Before propagating the erroneous derivatives back through the network, gradient clipping entails capping them. Smaller weights are the consequence of updating the weights using capped gradients. Clipping can be done on the values of gradients or on the norm of gradients. Both the clip value and clip norm options are available in optimizers like ADAM.

Adaptive Gradient Clipping (AGC): In [61] authors discussed that clipped gradient also converges faster than non-clipped gradients for general nonconvex problems. AGC [3] improves the convergence of gradient clipping by selecting an adaptive learning rate inversely proportional to the gradient norm, and ignoring the gradient's scale thus facilitating training with large batch sizes and strong data augmentations. AGC technique is also suggested as a replacement of Batch Norm process which enhances memory utilization and increases learning efficiency of Deep Neural Nets. Algorithm 1 describes the AGC process.

IV. Proposed Model and Methodology

For the purpose of image classification Normalization methods can be combined with popular convolutional neural network (VGG16, ResNet50, and Inception V3). To achieve this, the CT kidney dataset is subjected to nine different models.

Three different setups for CT Kidney image classification are build using three different deep neural nets i.e. VGG16, ResNet50, and InceptionV3 in each setup. First setup is build using Min Max normalization with Batch Norm technique. In second setup the Clip Value (CV) filter of ADAM optimizer is used, and in the third setup Adaptive Gradient Clipping (AGC) is used as a replacement of Batch Norm. Fig. 4 shows the complete experimental setup.

A. Dataset Description

The CT Kidney dataset is a collection of 12,446 distinct jpeg images of which 3,709 are related to cysts, 5,077 to normal, 1,377 to stones, and 2,283 to tumors. The Picture Archiving and Communication System (PACS) was used to collect the dataset of patients who had previously been diagnosed with kidney tumors, cysts, normal findings, or stones at different hospitals in Dhaka, Bangladesh. The initial collection was of DICOM (Digital Imaging and Communications in Medicine) images which contain multiple monochrome images along with patient information and other meta data. These images were converted to lossless JPEG image format and the patient information and meta data were removed. Few sample images from the four classes of the dataset are shown in Fig. 5, 6, 7, and 8.

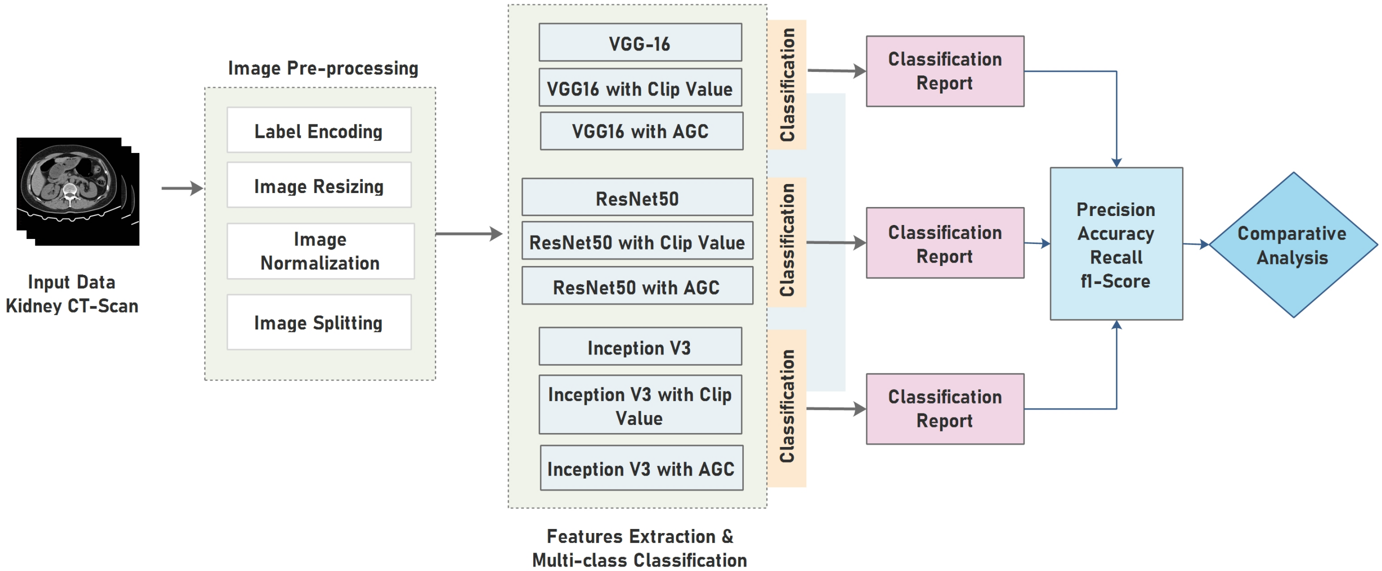

Fig. 4. Proposed model.

Fig. 5. Sample images for normal class of CT kidney dataset.

Fig. 6. Sample images for cyst class of CT kidney dataset.

B. Dataset Preprocessing

Images were initially resized to 150 by 150 pixels. The dataset is then normalized using min max normalization. The models are evaluated using a scheme where 80% of the images were taken to train the model and 20% to test the data. This resulted in 9,960 training image set and 2,491 test image set spread across 4 labels.

C. Classification Performance Metrics

A multiclass confusion matrix is used to calculate all the classification performance metric. All the classification models applied on CT Kidney dataset generates a multiclass confusion matrix from which the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative(FN) values are calculated for each class.

The TP, TN, FP, FN values taken from confusion matrix help us calculate the following metrics tabulated in Table I.

TABLE I. CLASSIFICATION PERFORMANCE METRICS

S.No.	Metric	Formula
1.	Accuracy	{(TP+TN)}/{(TP+TN+FP+FN)}
2.	Recall/Sensitivity	{TP}/{(TP+FN)}
3.	Precision	{TP}/{(TP+FP)}
4.	F1-score	{(Recall*Precision)}/{(Recall+Precision)}

Fig. 7. Sample images for stones class of CT kidney dataset.

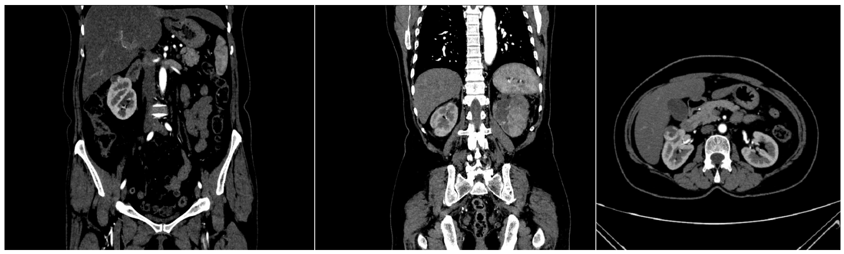

Fig. 8. Sample images for tumor class of CT kidney dataset.

Additionally for the multiclass dataset the Macro F1 score and the weighted F1 score are also calculated. The class-wise F1 scores acquired are simply averaged to get the macro-averaged F1 score of a model. The weighted F1 score is the average of the class-wise F1 scores, with the weights assigned based on the quantity of samples in each class.

Another powerful classification metric is Area Under Receiver Operating characteristics (AU-ROC) curve. The AU-ROC is also plotted but owing to the high accuracies achieved across the models AUROC score of 1 is achieved for all the classification models. Precision, recall, and f1-scores are used primarily for comparing the model's performance.

V. Results and Discussions

The experimental results of three distinct models are shown and discussed in this section. All three models used three different deep neural net VGG16, ResNet50, and InceptionV3 creating total of nine combinations. All the nine different combinations used the same hyperparameters for doing a fair comparison.

A. Experimental Setup

All the models were executed on Jupyter notebook, installed on Windows 10 platform with i9 (12th gen) processor running at 3.19 GHz and having 64 GB of RAM and 1 TB hard Disk Drive space.

The various Python libraries used were Keras, TensorFlow, Numpy, os, Sci-kit Learn, and Matplotlib. The hyperparameters are commonly used across all nine models and are tabulated in Table II.

Additionally for all the Clip Value (CV) based models clipvalue=0.6 is used and for all Adaptive Gradient Clipping (AGC) based models decay rate of 0.95 is used with initial Clip Norm value = 1.00. The clip norm value is updated adaptively by AGC algorithm during the run time.

TABLE II. Parameter Tuning in all Models

SNo	Parameter	Value
1	No. of Epocs	15
2	Learning Rate	0.01
3	Drop Out	0.5
4	Loss Function	Categorical Cross Entropy
5	Optimizer	ADAM
6	Batch Size	32
7	Activation Function	ReLU and Softmax
8	Regularization	Early Stopping from Keras
9	Preprocessing	Label Encoder/One Hot Encoding
10	Padding	'Same'

B. Performance Analysis of CT Kidney Dataset using VGG16

VGG16 is evaluated first on the CT Kidney dataset. The multi-class confusion matrix are shown in Fig. 9 and the classification metrics are displayed in Table III which displays the results of Plain VGG16, VGG16 with applied Clip Value (VGG16-CV), and VGG16 with Adaptive Gradient Clipping (VGG16-AGC).

In plain VGG16 classification accuracy of 96.9% is achieved which becomes 96.8% with CV model and 97.1% with application of AGC. The highest value of Macro and weighted f1-Score are also achieved with Adaptive Gradient Clipping technique i.e. 96.1% and 97.2% respectively. The highest values of precision are recorded in the plain model except for Class1 where the CV and AGC based models achieve higher precision values. Plain VGG16 achieves best Recall for Class0 and Class1 and AGC model achieves best Recall for Class1 and Class3.

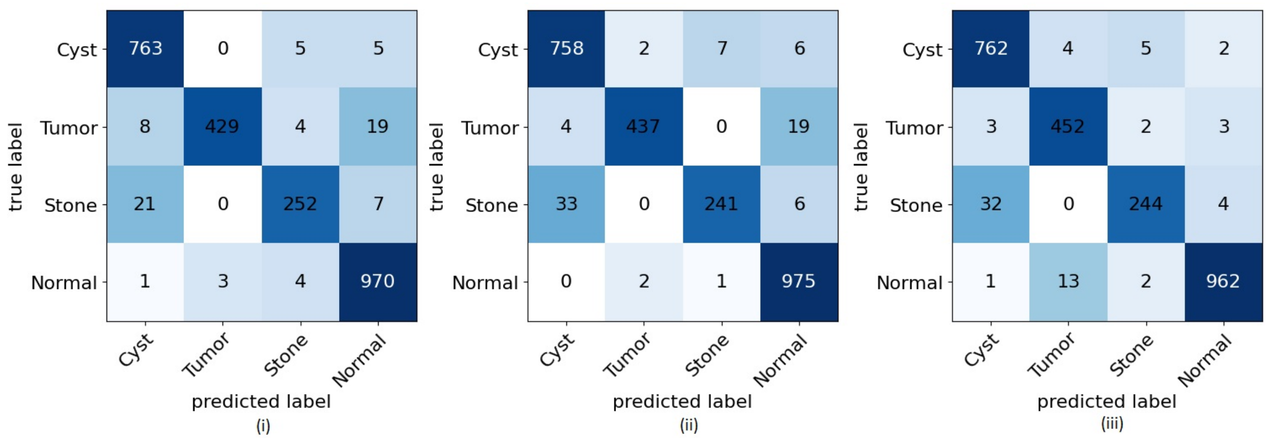

Fig. 9. Confusion Matrix for (i) Plain VGG16, (ii) VGG16-CV, (iii) VGG16-AGC.

TABLE III. Classification Report of VGG16 Models

	Plain VGG16			VGG16-CV			VGG16-AGC		
	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score
Class 0(Cyst)	.987	.962	.974	.981	.953	.967	.986	.955	.970
Class 1(Tumor)	.933	.993	.962	.950	.991	.970	.983	.964	.973
Class 2(Stone)	.900	.951	.925	.861	.968	.911	.871	.964	.916
Class 3(Normal)	.992	.969	.980	.997	.969	.969	.984	.991	.987
Accuracy		.969			.968			.971	
Misclassification		.031			.032			.029	
Macro f1		.960			.958			.961	
Weighted f1		.969			.968			.972	

C. Performance Analysis of CT Kidney Dataset using ResNet50

Next the performance of ResNet50 on the CT Kidney dataset is discussed. The three multi-class confusion matrix are shown in Fig. 10 and the classification metrics are displayed in Table IV, which contains the results of plain ResNet50, ResNet50 with applied Clip Value (ResNet50-CV), and ResNet50 with Adaptive Gradient Clipping (ResNet50-AGC).

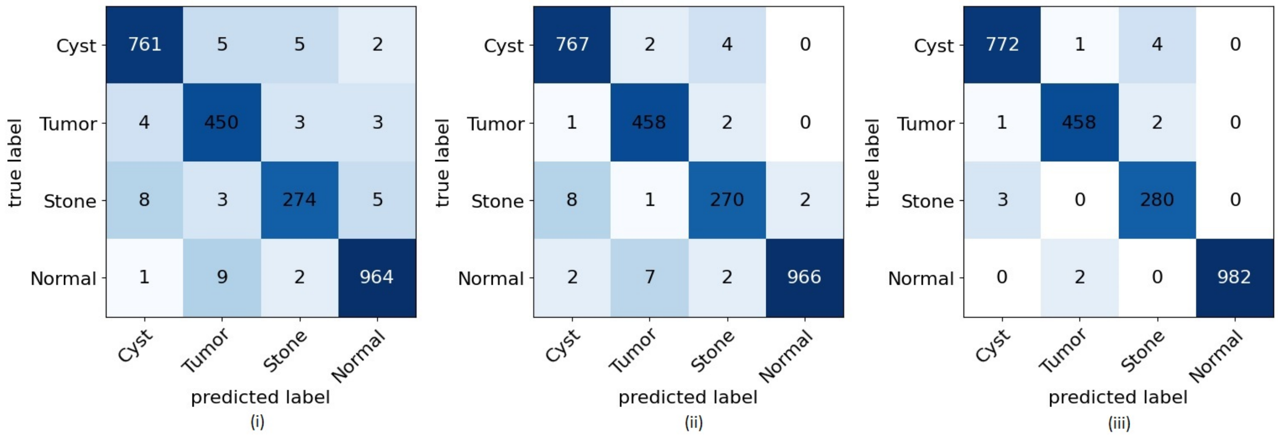

Fig. 10. Confusion Matrix for (i) Plain ResNet50, (ii) ResNet50-CV, (iii) ResNet50-AGC.

TABLE IV. Classification Report of ResNet50 Models

	Plain ResNet50			ResNet50-CV			ResNet50-AGC		
	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score
Class 0(Cyst)	.984	.983	.984	.992	.986	.989	.994	.995	.994
Class 1(Tumor)	.978	.964	.971	.993	.979	.986	.993	.993	.993
Class 2(Stone)	.945	.965	.955	.961	.971	.966	.989	.979	.984
Class 3(Normal)	.988	.990	.989	.989	.998	.993	.998	1	.999
Accuracy		.980			.988			.995	
Misclassification		.020			.012			.005	
Macro f1		.975			.984			.993	
Weighted f1		.980			.988			.995	

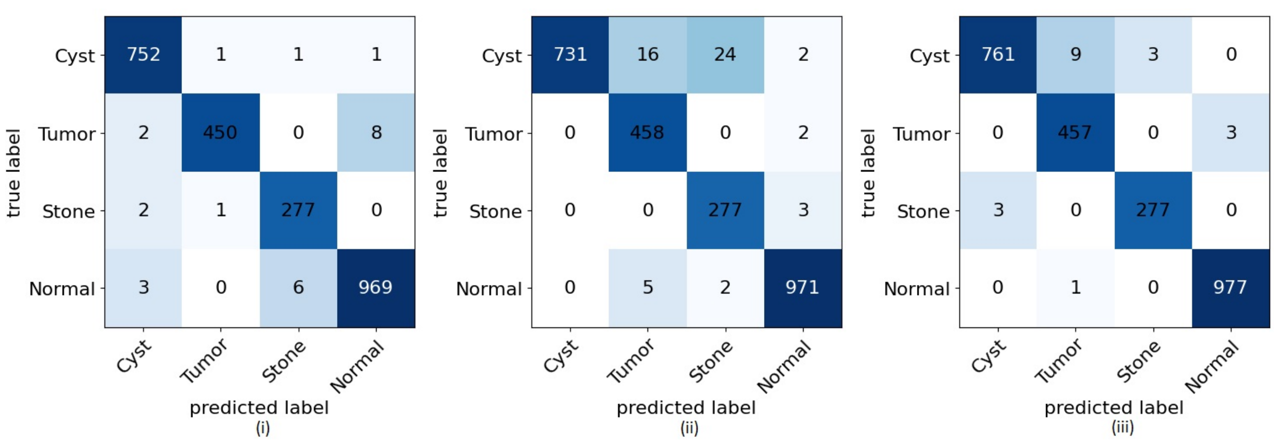

Fig. 11. Confusion Matrix for (i) Plain InceptionV3, (ii) InceptionV3-CV, (iii) InceptionV3-AGC.

TABLE V. CLASSIFICATION REPORT OF INCEPTION V3 MODELS

	Plain InceptionV3			InceptionV3-CV			InceptionV3-AGC		
	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score
Class 0(Cyst)	.996	.991	.993	.946	1	.972	.984	.996	.990
Class 1(Tumor)	.978	.996	.987	.996	.956	.976	.993	.979	.986
Class 2(Stone)	.989	.975	.982	.989	.914	.950	.989	.989	.989
Class 3(Normal)	.991	.991	.990	.993	.993	.993	.999	.997	.998
Accuracy		.990			.978			.992	
Misclassification		.010			.022			.008	
Macro f1		.988			.973			.991	
Weighted f1		.990			.978			.992	

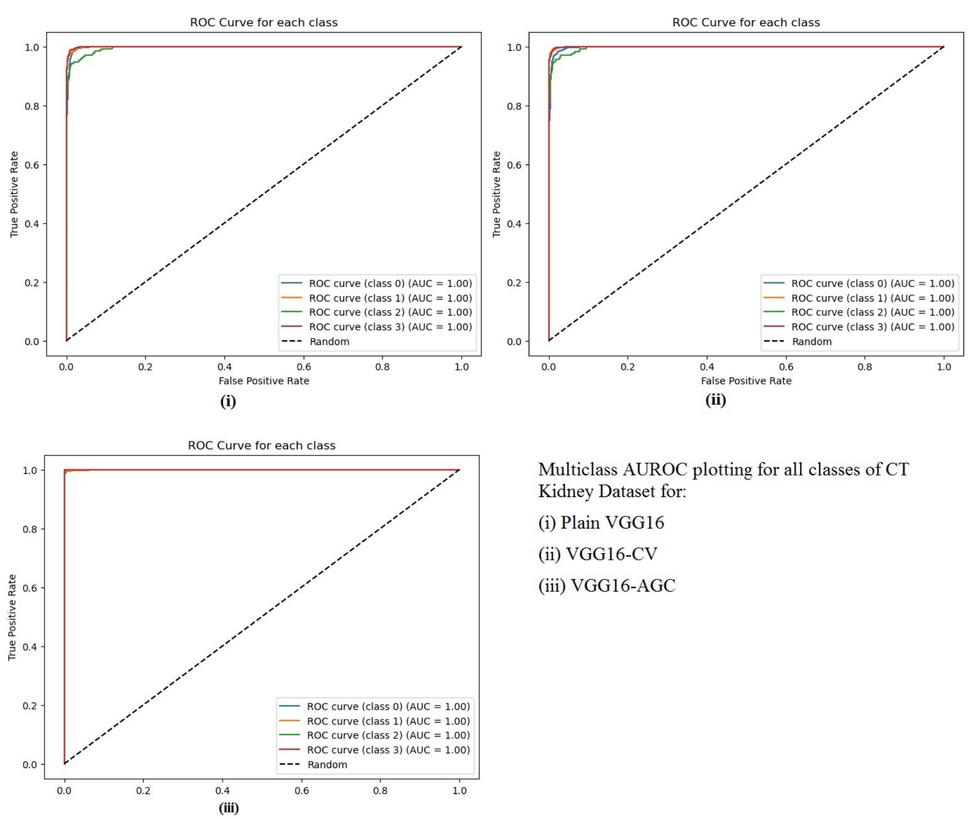

Fig. 12. All AUROC for VGG16 Setup.

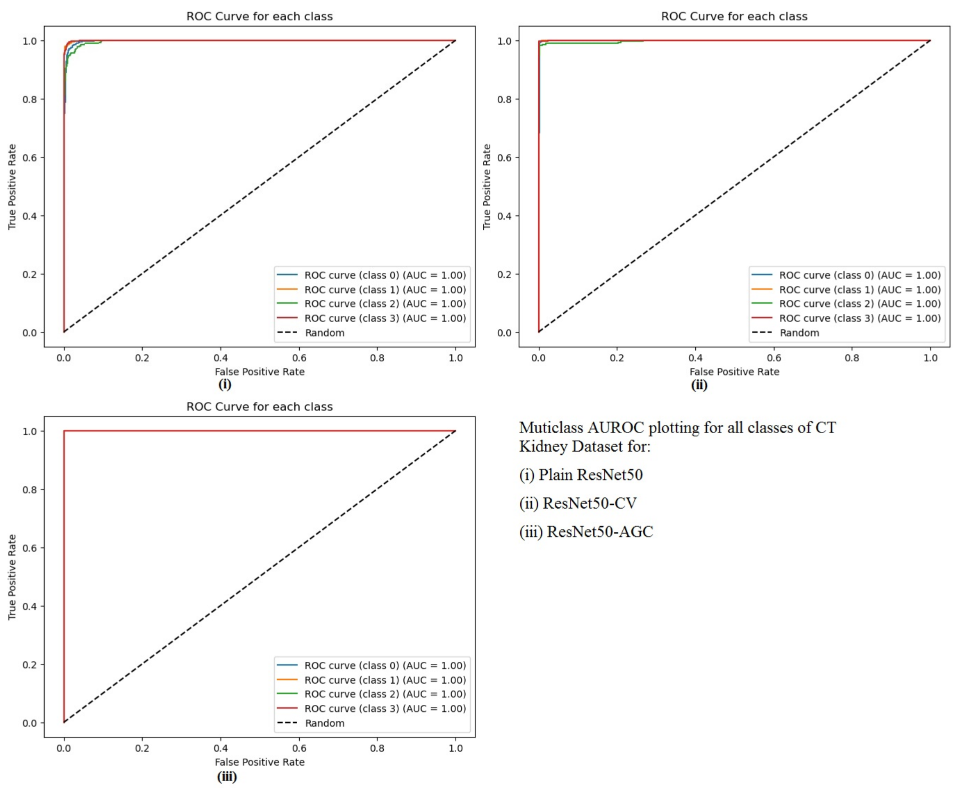

Fig. 13. All AUROC for ResNet50 Setup.

Plain ResNet50 achieved classification accuracy of 98% which increases to 98.8% with ResNet50-CV and 99.5% in ResNet50-AGC model. With AGC the highest scores of Macro F1, Weighted F1 are achieved. CV model produces marginally better results as compared to plain model but AGC model scores the highest precision, and recall values across all the classes giving the best results.

D. Performance Analysis of CT Kidney Dataset using Inception V3

In the last setup Inception V3 is used to build the multiclass classification models. The resultant confusion matrix are shown in Fig. 11 and the classification metrics are displayed in Table V for plain Inception V3, Inception V3 with applied Clip Value (Inception V3-CV), and Inception V3 with Adaptive Gradient Clipping (Inception V3-AGC). Plain InceptionV3 achieved classification accuracy of 99% which decreased to 97.8% for the CV model and increased to 99.2% with AGC model. The Clip Value model performs marginally better for Class3 but in the remaining classes it stays behind the plain model. AGC model excels in precision in Class1 and Class3 and in Recall in Class0 and Class3. Overall AGC model marginally stays ahead of Plain model with highest score of Macro F1, and Weighted F1.

The AUROC for all the nine models are also plotted in Fig. 12, 13, and 14. The AUROC score remains same across all models because of high accuracy achieved.

VI. Summary and Future Work

Two different techniques of Clip Value(CV) and Adaptive Gradient Clipping(AGC) are put to test on the CT Kidney Dataset in this article and their performances are compared with the plain models of VGG16, ResNet50, and Inception V3 models. Using clip value feature of optimizer brings marginal advantages in the VGG16 and ResNet50 models but fails to create much notable improvements. But with AGC model, a notable improvement is observed with VGG16 and ResNet50 setup, and marginal improvement with Inception V3 setup. AGC not only improves classification metrics but also improves the learning rates of the models. It was noted that during the training the AGC models quickly crossed the 90% and the 95% accuracy marks as compared to the plain models and the CV based models which took more epochs to reach 90% and 95% accuracies.

These results can be compared with the results in [62] and [2] in which authors have build classifiers for CT Kidney dataset. In [62] authors achieve accuracy of 95.29%, 99.48% and 97.38% using the MobileNetV2, VGG16, and InceptionV3 deep neural nets. In article [2] authors achieve accuracies of 98.20% using VGG16, 73.80% using ResNet50, and 61.60% using Inception V3. The results produced by the AGC based model achieves higher value of accuracy in VGG16, ResNet50, and Inception V3 as compared to these papers (Table VI). AGC is a promising technique that can be further tested on much larger datasets like Diabetic Retinopathy dataset in which the parameters of batch size and learning rates can altered to see the effect of AGC on larger batch sizes and higher learning rates. Still the existing results have proven that AGC technique can be a helpful method for training image datasets where it is difficult or time consuming to decide the clipping threshold for regularizing the train dataset.

TABLE VI. Accuracy Comparison of the Proposed AGC based Classification Method

Research Contribution	Method	Accuracy
M. H. K. Mehedi et al., 2022, [62]	MobileNetV2	95.29%
	VGG16,	99.48%
	InceptionV3	97.38%
M. N. Islam et al., 2022, [2]	Resnet	73.80%
	VGG16	98.20%
	Inception v3	61.60%
AGC (Proposed)	Resnet	99.5%
	VGG16	97.1%
	Inception v3	99.2%

References

[1] K. P. Andriole, J. M. Wolfe, R. Khorasani, S. T. Treves, D. J. Getty, F. L. Jacobson, M. L. Steigner, J. J. Pan, A. Sitek, and S. E. Seltzer, "Optimizing analysis, visualization, and navigation of large image data sets: one 5000-section ct scan can ruin your whole day," *Radiology*, vol. 259, no. 2, pp. 346–362, 2011.

[2] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from ct-radiography," *Scientific Reports*, vol. 12, no. 1, p. 11440, 2022.

[3] A. Brock, S. De, S. L. Smith, and K. Simonyan, "High-performance large-scale image recognition without normalization," in *International Conference on Machine Learning*. PMLR, 2021, pp. 1059–1071.

[4] J. Sundström, J. Bodegard, A. Bollmann, M. G. Vervloet, P. B. Mark, A. Karasik, T. Taveira-Gomes, M. Botana, K. I. Birkeland, M. Thuresson *et al.*, "Prevalence, outcomes, and cost of chronic kidney disease in a contemporary population of 2· 4 million patients from 11 countries: The careme ckd study," *The Lancet Regional Health–Europe*, vol. 20, 2022.

[5] C. P. Kovesdy, "Epidemiology of chronic kidney disease: an update 2022," *Kidney International Supplements*, vol. 12, no. 1, pp. 7–11, 2022.

[6] S. AC, "Chronic kidney disease: a public health problem that needs a public health action plan," *Prev Chronic Dis (Serial Online)*, vol. 3, p. 2, 2006.

[7] S. Jacobson, "Chronic kidney disease–a public health problem?" *Lakartidningen*, vol. 110, no. 21, pp. 1018–1020, 2013.

[8] J. B. Saaddine, K. V. Narayan, and F. Vinicor, "Vision loss: a public health problem?" *Ophthalmology*, vol. 110, no. 2, pp. 253–254, 2003.

[9] A. S. Levey, J. Coresh, K. Bolton, B. Culleton, K. S. Harvey, T. A. Ikizler, C. A. Johnson, A. Kausz, P. L. Kimmel, J. Kusek *et al.*, "K/doqi clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification," *American Journal of Kidney Diseases*, vol. 39, no. 2 SUPPL. 1, pp. i–ii+, 2002.

[10] S. Jacobson, "[chronic kidney disease–a public health problem?]," *Lakartidningen*, vol. 110, no. 21, p. 1018â€"1020, 2013. [Online]. Available: http://europepmc.org/abstract/MED/23805764

[11] S. M. Sozio, K. A. Pivert, F. J. Caskey, and A. Levin, "The state of the global nephrology workforce: A joint asn–era-edta–isn investigation," *Kidney international*, vol. 100, no. 5, pp. 995–1000, 2021.

[12] T. Gunasekara, P. M. C. De Silva, E. Ekanayake, W. Thakshila, R. Pinipa, P. Sandamini, S. Gunarathna, E. Chandana, S. Jayasinghe, C. Herath *et al.*, "Urinary biomarkers indicate pediatric renal injury among rural farming communities in sri lanka," *Scientific Reports*, vol. 12, no. 1, p. 8040, 2022.

[13] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, "A review on deep learning in medical image analysis," *International Journal of Multimedia Information Retrieval*, vol. 11, no. 1, pp. 19–38, 2022.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

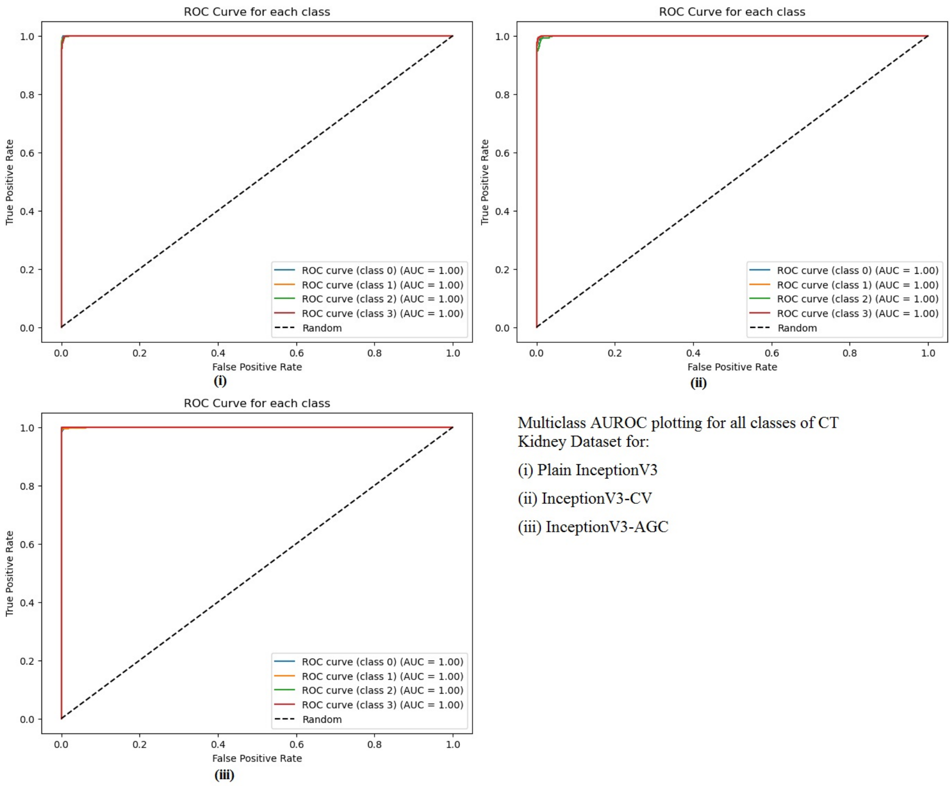

Fig. 14. All AUROC for InceptionV3 Setup.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[16] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[17] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[18] K. He, R. Girshick, and P. Dollár, "Rethinking imagenet pre-training," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4918–4927.

[19] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[20] O. Parkhi, A. Vedaldi, C. Jawahar, and A. Zisserman, "Cats and dogs, in 'computer vision and pattern recognition (cvpr)'," in *Computer Vision and Pattern Recognition (CVPR)*, 2012.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2023.

[22] K. Wang, C. Xu, G. Li, Y. Zhang, Y. Zheng, and C. Sun, "Combining convolutional neural networks and self-attention for fundus diseases identification," *Scientific Reports*, vol. 13, no. 1, p. 76, 2023.

[23] X. Pan, C. Ge, R. Lu, S. Song, G. Chen, Z. Huang, and G. Huang, "On the integration of self-attention and convolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 815–825.

[24] J.-B. Cordonnier, A. Loukas, and M. Jaggi, "On the relationship between self-attention and convolutional layers," *arXiv preprint arXiv:1911.03584*, 2019.

[25] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," *Advances in neural information processing systems*, vol. 32, 2019.

[26] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7794–7803.

[27] D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. Van Der Maaten, "Exploring the limits of weakly supervised pretraining," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 181–196.

[28] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves imagenet classification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 687–10 698.

[29] J. Cheng, S. Tian, L. Yu, C. Gao, X. Kang, X. Ma, W. Wu, S. Liu, and H. Lu, "Resganet: Residual group attention network for medical image classification and segmentation," *Medical Image Analysis*, vol. 76, p. 102313, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841521003583

[30] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby, "Big transfer (bit): General visual representation learning," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 2020, pp. 491–507.

[31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[32] N. Sharma, L. Malviya, A. Jadhav, and P. Lalwani, "A hybrid deep neural net learning model for predicting coronary heart disease using randomized search cross-validation optimization," *Decision Analytics Journal*, vol. 9, p. 100331, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772662223001716

[33] O. Moutik, H. Sekkat, S. Tigani, A. Chehri, R. Saadane, T. A. Tchakoucht, and A. Paul, "Convolutional neural networks or vision transformers: Who will win the race for action recognitions in visual data?" *Sensors*, vol. 23, no. 2, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/2/734

[34] E. M. Dogo, O. Afolabi, N. Nwulu, B. Twala, and C. Aigbavboa, "A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks," in *2018 international conference on computational techniques, electronics and mechanical systems (CTEMS)*. IEEE, 2018, pp. 92–99.

[35] R. Poojary and A. Pai, "Comparative study of model optimization techniques in fine-tuned cnn models," in *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. IEEE, 2019, pp. 1–4.

[36] J. Li, S. Fong, L.-s. Liu, N. Dey, A. S. Ashour, and L. Moraru, "Dual feature selection and rebalancing strategy using metaheuristic optimization algorithms in x-ray image datasets," *Multimedia Tools and Applications*, vol. 78, pp. 20 913–20 933, 2019.

[37] M. Awais, F. Shamshad, and S.-H. Bae, "Towards an adversarially robust normalization approach," *arXiv preprint arXiv:2006.11007*, 2020.

[38] S. D. Pande and R. Agarwal, "Multi-class kidney abnormalities detecting novel system through computed tomography," *IEEE Access*, vol. 12, pp. 21 147–21 155, 2024.

[39] N. Sharma and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of medical physics/Association of Medical Physicists of India*, vol. 35, no. 1, p. 3, 2010.

[40] Y. Tan, "Chapter 11 - applications," in *Gpu-Based Parallel Implementation of Swarm Intelligence Algorithms*, Y. Tan, Ed. Morgan Kaufmann, 2016, pp. 167–177. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012809362750011X

[41] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift fÃ¼r Medizinische Physik*, vol. 29, no. 2, pp. 86–101, 2019, special Issue: Deep Learning in Medical Physics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S093938891830120X

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.

[45] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[46] H. Zhang, Y. N. Dauphin, and T. Ma, "Fixup initialization: Residual learning without normalization," 2019.

[47] S. De and S. Smith, "Batch normalization biases residual blocks towards the identity function in deep networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 964–19 975, 2020.

[48] A. Brock, S. De, and S. L. Smith, "Characterizing signal propagation to close the performance gap in unnormalized resnets," *arXiv preprint arXiv:2101.08692*, 2021.

[49] P. Luo, X. Wang, W. Shao, and Z. Peng, "Towards understanding regularization in batch normalization," *arXiv preprint arXiv:1809.00846*, 2018.

[50] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, "Accurate, large minibatch sgd: Training imagenet in 1 hour," *arXiv preprint arXiv:1706.02677*, 2017.

[51] S. Singh and A. Shrivastava, "Evalnorm: Estimating batch normalization statistics for evaluation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3633–3641.

[52] C. Summers and M. J. Dinneen, "Four things everyone should know to improve batch normalization," *arXiv preprint arXiv:1906.03548*, 2019.

[53] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[54] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.

[55] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of tricks for image classification with convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 558–567.

[56] S. Shen, Z. Yao, A. Gholami, M. Mahoney, and K. Keutzer, "Powernorm: Rethinking batch normalization in transformers," in *International Conference on Machine Learning*. PMLR, 2020, pp. 8741–8751.

[57] Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[58] J. Shao, K. Hu, C. Wang, X. Xue, and B. Raj, "Is normalization indispensable for training deep neural network?" *Advances in Neural Information Processing Systems*, vol. 33, pp. 13 434–13 444, 2020.

[59] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*. Pmlr, 2013, pp. 1310–1318.

[60] S. Merity, N. S. Keskar, and R. Socher, "Regularizing and optimizing lstm language models," *arXiv preprint arXiv:1708.02182*, 2017.

[61] J. Zhang, T. He, S. Sra, and A. Jadbabaie, "Why gradient clipping accelerates training: A theoretical justification for adaptivity," *arXiv preprint arXiv:1905.11881*, 2019.

[62] M. H. K. Mehedi, E. Haque, S. Y. Radin, M. A. Ur Rahman, M. T. Reza, and M. G. R. Alam, "Kidney tumor segmentation and classification using deep neural network on ct images," in *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2022, pp. 1–7.