Comparision based Prediction of Diabetic Nephropathy using Deep Learning Algorithm

Saxena Sachin Kumar
Research Scholar
Department of Computer Science and Engineering
Faculty of Engineering & Technology
Invertis University
Bareilly, Uttar Pradesh - India

Shrivastava Jitendra Nath
Professor
Department of Computer Science and Engineering
Faculty of Engineering & Technology
Invertis University
Bareilly, Uttar Pradesh - India

Agarwal Gaurav
Associate Professor
Department of Computer Science and Engineering
Faculty of Engineering & Technology
Invertis University
Bareilly, Uttar Pradesh - India

ABSTRACT
India is the leading country in statistics in terms of mortality due to challenging hospital facilities and financial resources reach out to general people. Post COVID19, it has been researched that there is a big gap between traditional and recent medical facilities to patients inside hospitals all across the country. Furthermore, due to a lack of proper follow-ups and treatment of certain diseases like diabetes, millions of people are in extremis each day. This paper summarizes, the prediction of diabetic nephropathy of any patient using the deep learning image processing method namely VGG16 with 98 % accuracy. To accelerate the image training system design ROC and AUC curves have been defined also, to provide better results, optimal values have been compared with machine learning algorithms such as SVM, Random Forest, AdaBoost, etc. Patient images can be scanned digitally and the very first opinion can be obtained without expert knowledge acquisition. Dataset has been collected from Shri Ram Murti Smarak Hospital, Bareilly MRI, and Mission Hospital, Bareilly, Uttar Pradesh, India.

General Term
Deep Learning, Machine Learning

Keywords
Deep learning, machine learning, medical care, diabetic nephropathy, VGG16

1. INTRODUCTION
According to the International Diabetes Federation, over 460 million people are suffering from diabetes disease. Moreover, it has become one out of 10 causes of global deaths all across the world. Post COVID19, traditional and new ways of medical care are also influencing the growth rate of this disease. [1] The risk of early death from diabetes is also increasing day by day, according to the recent report of the World Health Organization. Up to the present research, it has been observed that most COVID19 patients were suffering from diabetes. Deaths brought about by Diabetes can be predicted earlier by the hospital management as well. More than 422 million people worldwide have diabetes, significantly in low-and middle-income countries, and 1.6 million deaths are directly attributed to diabetes each year. In India, the pie chart is shown below. [13] With the hazardous result of this disease, most active parts of the body such as the heart, blood vessels, eyes, kidneys, and nerves can be damaged. [2] India had observed 1,150300 deaths in 2019 with an age range of 20-70 years, which is among the top five countries among Bangladesh, Nepal, Sri Lanka, and Mauritius. Moreover, it is assuming that around 50 million are still not diagnosed with diabetes yet. In the upcoming year, the graph will increase unexpectedly [3].

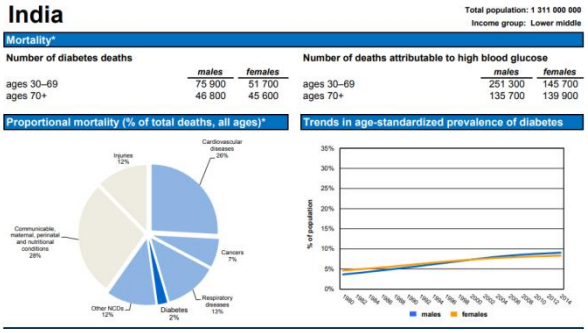

Figure 1 Statistics showing India's Diabetic deaths in India

2. MOTIVATION
A recent piece of research provides that as per primary diagnosis, 44.2 percent of the cause of kidney failure is diabetes. Diabetic nephropathy or diabetic kidney disease determines the symptoms, causes and treatment of kidneys. Diabetic nephropathy can analysis corresponds to urinary excretion, secretion of proteins, chronic hyperglycemia, malignant presence in the kidney etc.

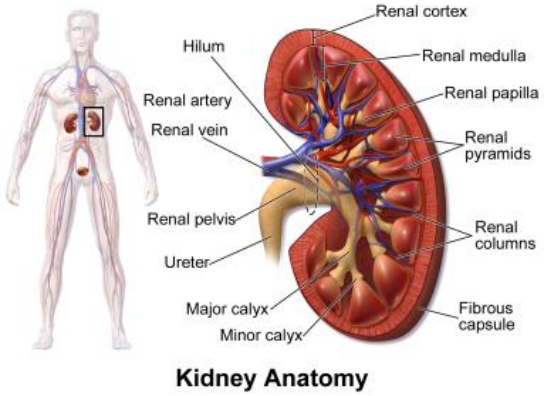

Figure 2 most body functionalities are performed by kidney

2.1. Types of Diabetic nephropathy
Kidney nephropathy is the study of mellitus functionality inside stomach of human body. There are excess of glucose and insulin in body hence resulting diabetic causes mainly of two types, type 1 and type 2. For patients suffering from type 1, four percent will develop diabetic nephropathy within ten years, and

twenty-five percent will develop diabetic nephropathy within twenty-five years. Moreover, For patients suffering from type 2, ten percent will develop diabetic nephropathy within five years, and thirty percent will develop diabetic nephropathy within twenty years.

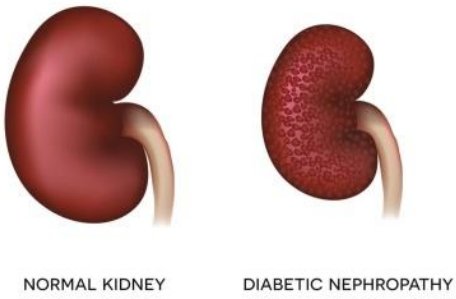

Figure 3 worst stage of Diabetic nephropathy

3. RELATED WORK

- The X-ray can go under training using a convolutional neural network to determined COVID19 or COVID19 patients through image augmentation. Residual block and ReLu function is accelerating the result to scan CT images of patients. In this research, great initiatives has been taken and images to be trained from pre-uploaded data on web resources, moreover, it is also defined that RT-PCR results are insufficient to classify. [8]

- The early detection of COVID19 patients using CT-scan images with the help of a deep learning algorithm namely DenseNet121. FJCovNet, a new kind of deep learning gives 98.14% accuracy within a time limit of 612 seconds only. Batch size of 64, 100 epochs, Adam optimizer, and 37 steps per epoch were sufficient to accelerate the result. Comparison with Xception, VGG16, Resnet50 with the proposed algorithm achieved satisfactory performance to diagnose the symptoms of COVID19. [9]

- Up to the present research, the Deep belief network also gives better results to diagnose chronic kidney disease. With an accuracy of 98.5%, the author has trained the proposed model to find out the four stages of chronic kidney disease and 25 characteristics of patients were recorded to performing training and testing data sets of UCI's machine learning. [10]

- The key aspects of utilization of deep learning in medical science are that CT images have been used in various research areas, to produce consistent quality to automatic detection of kidney lesions using morphological cascade convolutional network. The purpose of this examination and referral process is to differentiate six different sizes of lesions. ResNet101 is the part of research review to extract shallow features of CT-scan images. [11]

- In today's society, the medical care problem is more challenging when it associates with hardware or physical resources. Corresponds to heterogeneous medical science pieces of equipment, traditional and new ways of communication are difficult. So to make it easy, the research paper has summarized the Internet of things to predict chronic kidney disease using a Deep convolutional neural network. Abdominal CT-scan images are trained and validated to produce results using an adaptive hybridized deep convolutional neural network. [12]

- CT scan images are segmented by Submanifold Sparse Convolutional Networks. Kidney tumor segmentation aims to accurately in this paper. [14]

- At North South University, Bangladesh, the team was subjected to a Critical Patient Caring system in hospitals using Machine Learning's various methods. IBM Cloud as PaaS (Platform-as-a-Service), a cloud-based computing model that allows development teams to build, test, deploy, manage, update, and scale applications faster and more cost-effectively. Fascinating seven machine learning methods, they arrived at this result that IBM cloud can bring 92% of accuracy was previously 80-90% only. [15]

- Author from the Wuhan University of Technology Wuhan, China, focused on tumor treatment in the health care system. To demonstrate the results, the author used Artificial intelligence and machine learning methods, prominently SVM(Support Vector Machine) method. Furthermore, he narrowed down the research to Breast Cancer prediction using Machine Learning. [16]

- The objective of this study was to design and develop a 30-day risk of hospital readmission predictive model using machine-learning techniques. The study consisted of 180,118 admissions with 22565 (12.5%) of actual readmissions within 30-days of hospital discharge, from 01 Jan 2015 to 31 Dec 2016 from two Auckland-region hospitals. They have proposed a new model to predict various parameters using Machine Learning. [17]

4. PROBLEM DEFINTION AND PROPOSED SOLUTION

Three main types of Diabetes are existing mainly Type 1 diabetes which causes lack of insulin, type 2 diabetes causes ineffective use of insulin and type 3 Gestational diabetes because of hyperglycaemia with blood glucose values mostly pregnant women are suffering from.[4]

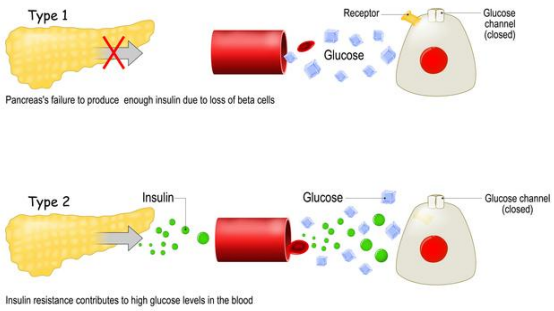

Figure 4 the effect of type 1 and type 2 on kidney

About 30 % of patients with Type 1 diabetes and 10 to 40 % of those with Type 2 diabetes are likely to suffer with kidney failure [5]. It has been observed that half of the world population directly or indirectly suffering from diabetes. Functionality level of kidneys also proportionally depends upon sugar level in body.

In today's society, medical care problem is alarming and patient treatment is really hard. Diagnose the type1 and type2 diabetics nephropathy using CT-scan, MRI and X-ray can save time with

better accuracy. Hence, this research helped to diagnose the category of diabetic nephropathy.

4.1 Deep Learning

Deep learning is a subset of machine learning, to solve any problem statement and uses neural network to simulate human like decision-making. Deep learning requires little guidance from the researchers, also extracts the features from given datasets whereas machine learning learns with datasets. It has in-built automatic multi-stage feature learning process; furthermore, there is evidence to suggest that it learns rich hierarchical representation. It is widely used in research because of best application of Artificial Neural Network. In this research paper, deep learning algorithms are deployed to classify type of diabetes based on extreme fine features of malignant in kidney. Deep learning algorithms such as VGG16, ResNet50, Inception Net, generative adversial network are needed huge annotated features which is very difficult to acquire. Machine learning and shallow neural network are not seen so effective in comparison to deep learning.[7] Working with Machine Learning and Deep Learning in Healthcare systems, patients captivating health issues and allude to lack of treatment can be reduced. The basic architecture of Deep learning is given in figure 1.

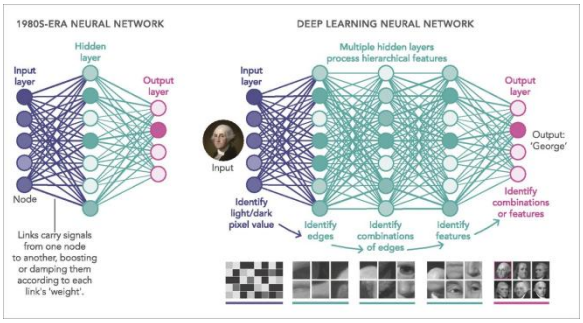

Figure 5 Deep Learning architecture

Images are pre-processed before feature extraction, then passes through several under conditions as depicted in fig 6.

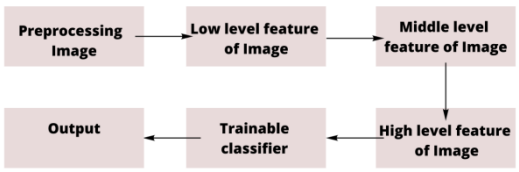

Figure 6 steps followed in deep learning

5. SYSTEM DESIGN

To provide a better result, we require other models too to compare with deep learning models. Following are the machine learning models to train and test the dataset collected from various hospitals.

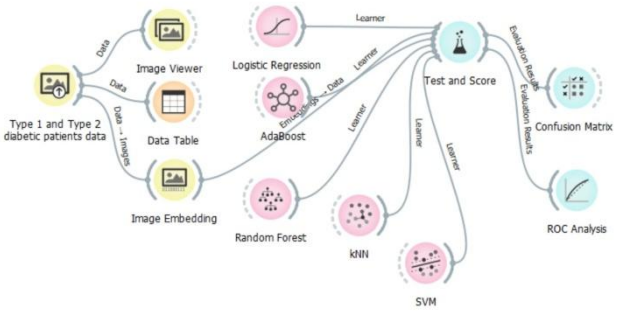

Figure 7 Machine learning algorithm training and testing phase

5.1 KNN model

This model is used as an application of a classification and regression problem. It finds out a similar feature object to produce better accuracy. In general, to calculate this distance, Euclidean distance is most widely used. Decision boundary also separates two classes. If D is the distance between two neighbours, then square root of difference between respective coordinate will be calculated as,

$$D(O_1,O_2)=\sqrt{(X_{11}-X_{21})^2+(X_{12}-X_{22})^2}$$

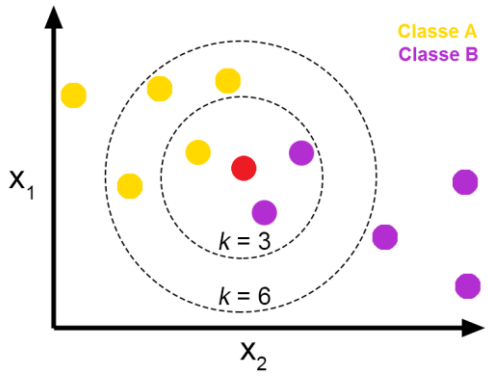

Figure 8 nearest neighbor cluster formed when k =3 and k=6 for Class A and Class B

5.2 SVM model

Machine learning algorithms such as support vector machines to search for optimal values out of several other essential hyperparameters. To classify two different features support vector machine defines a decision boundary with the nearest closest elements of these two classes. ROC and AUC curves can be utilized for finding the prediction model.

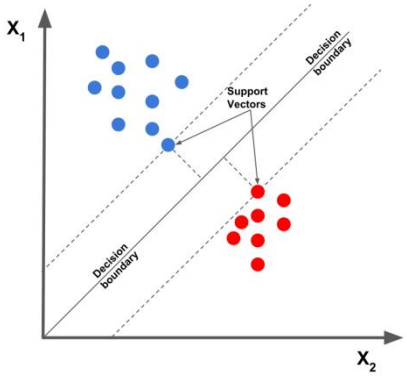

Figure 9 Two classes are separated by support vector and decision boundary

5.3 Random Forest

It is an ensemble technique, used in industries due to high scalability it includes the number of decision trees, number of records and features to be sampled, and Gini impurity index. It also uses a mean decrease in impurity to find the essential feature out of an entire tree.

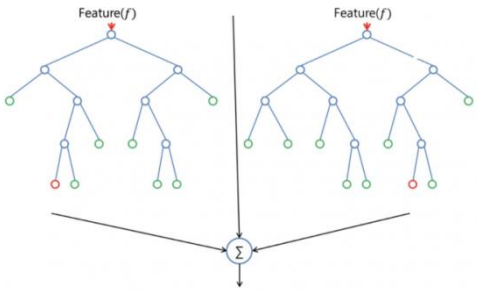

Figure 10 feature extracted from each tree to generate best output eventually

5.4 Logistic Regression

A probability-based model is used to classify binary values to obtain an optimal output of any model. There must be threshold value to determine the respective class category. The images are trained and tested using this model to find the category that belongs to type 1 diabetic or type 2 diabetic.

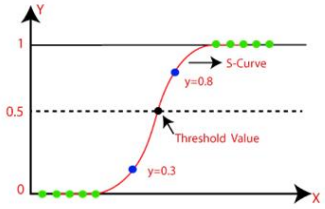

Figure 11 Logistic regression

5.5 AdaBoost

While training images, using the AdaBoost algorithm, a category of boosting algorithms, it assigns weighted values to each feature at the first iteration. This algorithm calculates the final sum of weighted all classifiers as a whole.

$$F(X_i)=\text{sign}\left(\sum \alpha_k f_k(X_i)\right)$$

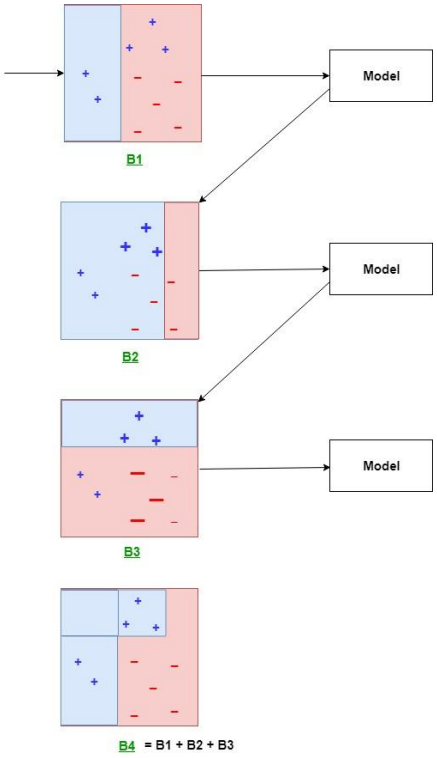

Figure 12 Adaboost algorithm process

Following figure shows the flow chart of image scanning process to determine the category of diabetic nephropathy.

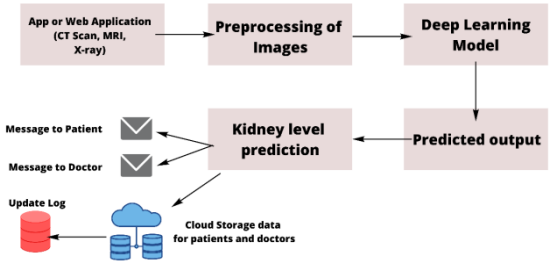

Figure 13 proposed system design of Deep learning model to train the image

5.6 VGG16

As image has very dense features, so this algorithm is used with 16 layers with weights and around 138 million parameters. It uses convolution and ReLu function to extract features, following architecture depicts 2 fully connected and other iterations involved in convolution and max layer.

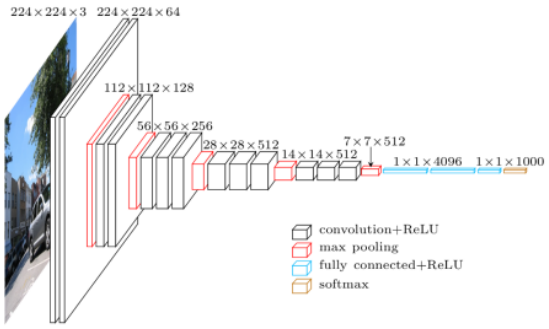

Figure 14 VGG16 architecture

In order to train and feature extraction following parameters has been taken as input in VGG16 algorithms.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 128, 128, 32)      416
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 128, 128, 32)      4128
_________________________________________________________________
batch_normalization (BatchNo (None, 128, 128, 32)      128
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 64, 64, 32)        0
_________________________________________________________________
dropout (Dropout)            (None, 64, 64, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 64, 64, 64)        8256
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 64, 64, 64)        16448
_________________________________________________________________
batch_normalization_1 (Batch (None, 64, 64, 64)        256
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 32, 32, 64)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 32, 32, 64)        0
_________________________________________________________________
flatten (Flatten)            (None, 65536)             0
_________________________________________________________________
dense (Dense)                (None, 512)               33554944
_________________________________________________________________
dropout_2 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 1026
=================================================================
Total params: 33,585,602
Trainable params: 33,585,410
Non-trainable params: 192
_________________________________________________________________
None
```

6. DATA COLLECTION

Around 53 patient's diabetes history recorded digitally [questionnaire link [6]] on the basis of age group, symptoms, urine reports, COVID19 occurrence etc., moreover their MRI, CT and X-ray images process to diagnose the type of diabetes. Blood filtration is most important function of kidneys. This research paper aims to diagnose type of diabetes in certain patient by analysis of their kidneys report. Significantly, higher in glucose level in body diabetic patients can damage blood vessels inside kidneys, body nervous system and urinary tract.

Table 1 consortium names and respective statistics of collected data

Data	Data collection source			
	Hospital name	***City***	***No of Images***	***Modality***
1.	Mission Hospital	Bareilly	20	X-Ray
2.	Bareilly MRI	Bareilly	8532	MRI, CT-scan
3.	SRMS Hospital	Bareilly	3044	MRI Images
		Total	11073	

None dataset has been used for training, testing and validation from other resources. To produce consistent quality of results, this research has taken MRI, CT, and X-ray modalities. The following are the features of the images taken as input :

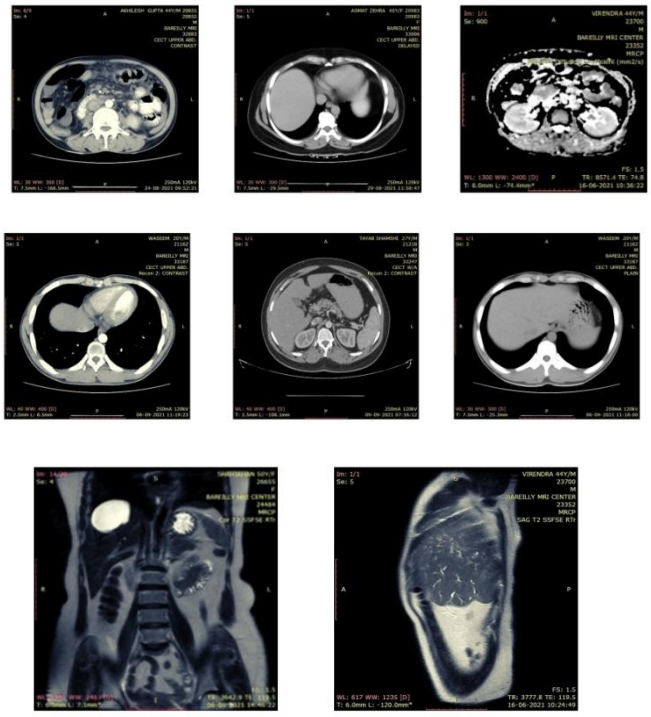

Figure 15 MRI images of kidney and stomach

Sample Patient's data:

SOP Class.....: 1.2.840.10008.5.1.4.1.1.2 (CT Image Storage)

Patient's Name...: KARAN SINGH 11Y/M,

Patient ID.....: 21173

Modality.....: CT

Study Date.....: 20210907

Image size.....: 550 x 888

Pixel Spacing....: [0.545455, 0.596847]

Slice location...: 0.000

7. RESULT: TRAINING AND TESTING OF MODELS

As deep learning is a feature extraction algorithm so following table represents the model performance of different models. VGG16 and other models are train with following set of training and testing sets.

Table 2 Teating and traing separte values

Data	Training and Testing set		
	Type 1	***Type 2***	
Training	1459	5908	7367
Testing	808	3421	4229
		Total	11596

Table 3 clearly mentioned the refined performance of classification of images with 0.987 AUC value and 0.983 CA value.

Table 3 Comparison based study after training

Model used	Model performance				
	AUC	***CA***	***F1***	***Precision***	***Recall***
KNN	0.984	0938	0.937	0.937	0.938
SVM	0.977	0.912	0.910	0.916	0.912
Random Forest	0.963	0.925	0.923	0.927	0.925
Logistic Regression	0.985	0.925	0.924	0.925	0.925
AdaBoost	0.833	0.875	0.872	0.875	0.854
VGG16	**0.987**	**0.983**	**.0988**	**0.982**	**0.994**

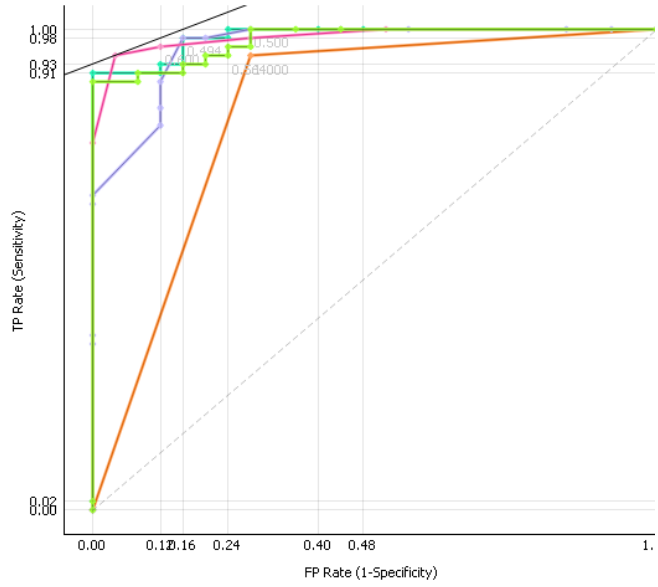

Figure 16 RUC curve

8. CONCLUSION

This paper summarizes some applications of deep learning to the prediction of extreme cases like cardiac attacks, chronic diseases of Diabetes Patients. A tiny contribution to "Ayushman Bharat National Protection Mission", an Indian government national-level project. In future the result can be improved by using more Deep Learning algorithms.Coordination between Doctors, Patients, and Health Care Policy. To put it another way, it aims to save money from expensive therapy. Diabetes mellitus patients can regulate themselves by an alert from various dimensions.

9. ACKNOWLEDGMENT

The first and foremost I acknowledge god to provide wisdom to research in ethical way. Then all the doctors, medical staff and supporting team of Shri Ram Murti Smarak Hospital, Bareilly MRI, and Mission Hospital, Bareilly, Uttar Pradesh, India, where data has been collected at regular interval, with this mind, I want to give thanks my guide Shrivastava Jitendra Nath, Professor and co-guide Agarwal Gaurav, Associate Professor, Department of Computer Science and Engineering, Faculty of Engineering & Technology, Invertis University, Bareilly, Uttar Pradesh – India.

10. REFERENCES

[1] [Online]. Available: https://www.idf.org/news/225:idf-ncda-statement-wha-resolution.html.

[2] [Online]. Available: https://www.who.int/news/item/14-04-2021-new-who-global-compact-to-speed-up-action-to-tackle-diabetes.

[3] [Online]. Available: https://idf.org/our-network/regions-members/south-east-asia/diabetes-in-sea.html.

[4] [Online]. Available: https://www3.paho.org/hq/index.php?option=com_content&view=article&id=6717:2012-about-diabetes&Itemid=39447&lang=en.

[5] [Online]. Available: https://www.kidney.org/atoz/content/diabetes.

[6] [Online]. Available: https://docs.google.com/forms/d/e/1FAIpQLScX3kpkPyBPDyCndl2_hWl27qwrkaDDlGZVEGLWoeWokAWV0A/alreadyresponded.

[7] "Y. Pan, M. Fu, B. Cheng, X. Tao and J. Guo, 'Enhanced Deep Learning As-sisted Convolutional Neural Network for Heart Disease Prediction on the In-ternet of Medical Things Platform,' in IEEE Access, vol. 8, pp. 189503-189512, 2020, doi: 10.1109/ACCESS.2020.3026214."

[8] "Z. KARHAN and F. AKAL, 'Covid-19 Classification Using Deep Learning in Chest X-Ray Images,' 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299315."

[9] "J. C. Sangidong, H. D. Purnomo and F. Y. Santoso, 'Application of Deep Learning for Early Detection of COVID-19 Using CT-Scan Images,' 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), 2021, pp. 61-65, doi: 10.1109/EIConCIT50028.2021.9431887."

[10] "S. M. M. Elkholy, A. Rezk and A. A. E. F. Saleh, 'Early Prediction of Chron-ic Kidney Disease using Deep Belief Network,' in IEEE Access, doi: 10.1109/ACCESS.2021.3114306."

[11] "H. Zhang, Y. Chen, Y. Song, Z. Xiong, Y. Yang and Q. M. Jonathan Wu, 'Automatic Kidney Lesion Detection for CT Images Using Morphological Cascade Convolutional Neural Networks,' in IEEE Access, vol. 7, pp. 83001-83011, 2019, doi: 10.1109/ACCESS.2019.2924207."

[12] "G. Chen et al., 'Prediction of Chronic Kidney Disease Using Adaptive Hy-bridized Deep Convolutional Neural Network on the Internet of Medical Things Platform,' in IEEE Access, vol. 8, pp. 100497-100508, 2020, doi: 10.1109/ACCESS.2020.2995310."

[13] [Online]. Available: https://www.who.int/health-topics/diabetes#tab=tab_1.

[14] "X. Yan, K. Yuan, W. Zhao, S. Wang, Z. Li and S. Cui, 'An Efficient Hybrid Model for Kidney Tumor Segmentation in CT Images,' 2020 IEEE 17th Inter-national Symposium on Biomedical Imaging (ISBI), 2020, pp. 333-336, doi: 10.1109/ISBI45749.2020.9098325."

[15] "A. A. Neloy, S. Alam, R. A. Bindu and N. J. Moni, 'Machine Learning based Health Prediction System using IBM Cloud as PaaS,' 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 444-450, doi: 10.1109/ICOEI.2019.8862754."

[16] "H. Dou, 'Applications of Machine Learning in The Field of Medical Care,' 2019 34rd Youth Academic Annual

Conference of Chinese Association of Au-tomation (YAC), 2019, pp. 176-179, doi: 10.1109/YAC.2019.8787685."

[17] "M. M. Baig et al., 'Machine Learning-based Risk of Hospital Readmissions: Predicting Acute Readmissions within 30 Days of Discharge,' 2019 41st An-nual International Conference of the IEEE Engineering in Medicine and Biolo-gy Society (EMBC), 2019, pp. 2178-2181, doi: 10.1109/EMBC.2019.8856646."

[18] Kakde, A., Arora, N., & Sharma, D. (2019). A comparative study of different types of cnn and highway cnn techniques. Global Journal of Engineering Science and Research Management, 6(4), 18-31.

[19] Kakde, A., Arora, N., Sharma, D., & Sharma, S. C. (2020). Multi spectral classification and recognition of breast cancer and pneumonia. Polish Journal of Medical Physics and Engineering, 26(1), 1-9.

[20] Kakde, A., Sharma, D., Kaushik, B., & Arora, N. (2021). Investigation of Solar Flare Classification to Identify Optimal Performance. ELCVIA Electronic Letters on Computer Vision and Image Analysis, 20(1), 28-41.

[21] Kakde, A., Arora, N., & Sharma, D. (2018). Fire Detection System Using Artificial Intelligence Techniques. International Journal of Research in Engineering, IT and Social Sciences, 8(11), 1-5.