VOL. 17, NO. 23, DECEMBER 2022 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

MULTI-CLASS CLASSIFICATION OF PROSTATE CANCER MRI BASED ON UCLA SCORE USING DEEP TRANSFER LEARNING MODELS

Rajesh M. N. and Chandrasekar B. S.
Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, Jain (Deemed-to-be University), Bangalore, Karnataka, India
E-Mail: rajeshmn.mn@gmail.com

ABSTRACT

Prostate cancer is the most frequent cancer in the men population, and early detection is critical in order to lower mortality rates from the disease. With its superior soft-tissue contrast, magnetic resonance imaging (MRI) has become the imaging method of choice for the localization of PCa. In terms of diagnosing PCa of the transition zone, T2-weighted images are the most useful tool among the several MRI modalities. In this proposed model, the PCa is classified based on the T2w MRI data. The proposed model is a deep learning approach, which includes the deep transfer learning models for the classification of PCa. For classifying the data, the different variants of VGG-16, VGG-19, and MobileNet-v3Large transfer learning models are used. These models are modified using different optimizers for varying the learning rate. Optimizers like Adam, AdaMax, SGD, RMSprop, and Ftrl are used in this research. For evaluation, the ampMRI dataset with 845 patient records with unique "UCLA" scores of the ROI was used for multi-class classification. For performance analysis, accuracy, sensitivity, specificity, precision, and F1 score are computed based on the classification. Finally, according to the results, the performance was compared among the different proposed models for validation. The proposed models optimized using the Ftrl optimizer have obtained better performances with 93.31% accuracy, 93.92% accuracy, and 95.27% accuracy for VGG-16-Model-04, VGG-19-Model-04, and MobileNet-v3 respectively.

Keywords: prostate cancer, VGG-16, VGG-19, mobileNet-v3, deep learning, prostate classification, T2w MRI.

I. INTRODUCTION

Prostate cancer (PCa) is the most frequent cancer in men today, with one in every six men developing the disease at a certain stage in their life. It has been shown that early identification and precise diagnosis of cancer could enhance the survival of cancer while also lowering treatment costs [1]. According to the Global Cancer Statistics 2020 report, PCa is the second leading cancer among men in the globe, following lung cancer. PCa, in contrast to lung cancer, which affects 14.3 percent of the world's population, affects 14.1 percent of men globally and has a death rate of 6.8 percent, according to the World Health Organization [2].

Earlier, TRUS (transrectal ultrasound) was the primary imaging tool used to assess individuals suspected of having PCa. However, this technique has a number of disadvantages, including low sensitivity and specificity rates, which are particularly prevalent in the lesions of the transition zone. More recently, magnetic resonance imaging has proven improved accuracy in diagnosing and is becoming a medical standard of care for individuals at risk of developing PCa [3]. In recent years, since the emergence of multi-parametric MRI (mpMRI), the imaging of PCa has made significant progress. It is possible that computer-aided quantitative analysis of prostate MRI will enhance PCa detection and will aid in the mpMRI interpretation standardization. Due to the superior soft-tissue contrast provided by MR imaging, it turned out as the imaging method of selection for the localization of PCa. In addition to higher-resolution T2-weighted (T2W), the mpMRI provides dynamic contrast-enhanced imaging (DCE-MR), diffusion-weighted imaging (DWI), and magnetic resonance spectroscopy (MRS). The use of mpMRI has been shown to be an effective approach for localizing high-risk PCa [4].

In general, there are four major MRI imaging modalities that are utilized in diagnosing PCa. T2W, DCE, DWI, and MRS are some of the techniques available. T2w-MRI is the most fundamental MRI imaging modality, and it works by constructing the grayscale images of the scanned item by utilizing the transverse relaxation time T2. T2w-MR images have proven to be an important imaging modality for noninvasive PCa diagnosis as their popularity and availability among healthcare professionals have grown in recent years. Using this modality, the primary advantage is that it enables visual differentiation between normal prostatic and malignant tissues by varying the homogeneity and intensity of the images. Its second major advantage is that it computes the data on the zonal structure of the prostatic gland, which is another major advantage [5].

T2w images are the most effective way to visualize the anatomy of the prostate gland. mp-MR imaging protocol, the capture of higher-resolution T2-weighted images of the prostate is the first and most significant stage in the procedure. When it comes to detecting PCa in the transition zone, T2w magnetic resonance imaging has been touted as an accurate approach [6]. T2w MR images are very useful in predicting the pathological stage of PCa as well as the extent of extracapsular expansion. Because T2w MR images are critical in the diagnosis and stage grading of PCa, T2w is the foundation and most crucial sequence in computer-aided diagnosis and staging systems for PCa. This research proposed the utilization of deep transfer learning (DTL) models based on deep learning architecture to classify PCa using T2w-MRI images. In

T2w images, the cancer region of interest (ROI) has many darker pixels than the brighter pixels, whereas the normal tissue ROI has many brighter pixels than darker pixels [7].

For this work, the deep transfer learning (DTL) technique was chosen above other machine-learning algorithms because of the remarkable classification accuracy of pre-trained models. Additionally, the pre-trained model saves time consumption by evading the issue of retraining and testing the approach weights from the beginning of the training process. Using T2w-MRI images, this work recommends the use of DTL models such as VGG-16, VGG-19, and MobileNet-v3 for the classification of PCa. Multiple alternative optimizers, such as Adam, AdaMax, SGD, RMSprop, and Ftrl, are used to train each of these DTL models, with appropriate learning rates, in order to reduce the generalization factor, which is an important concern.

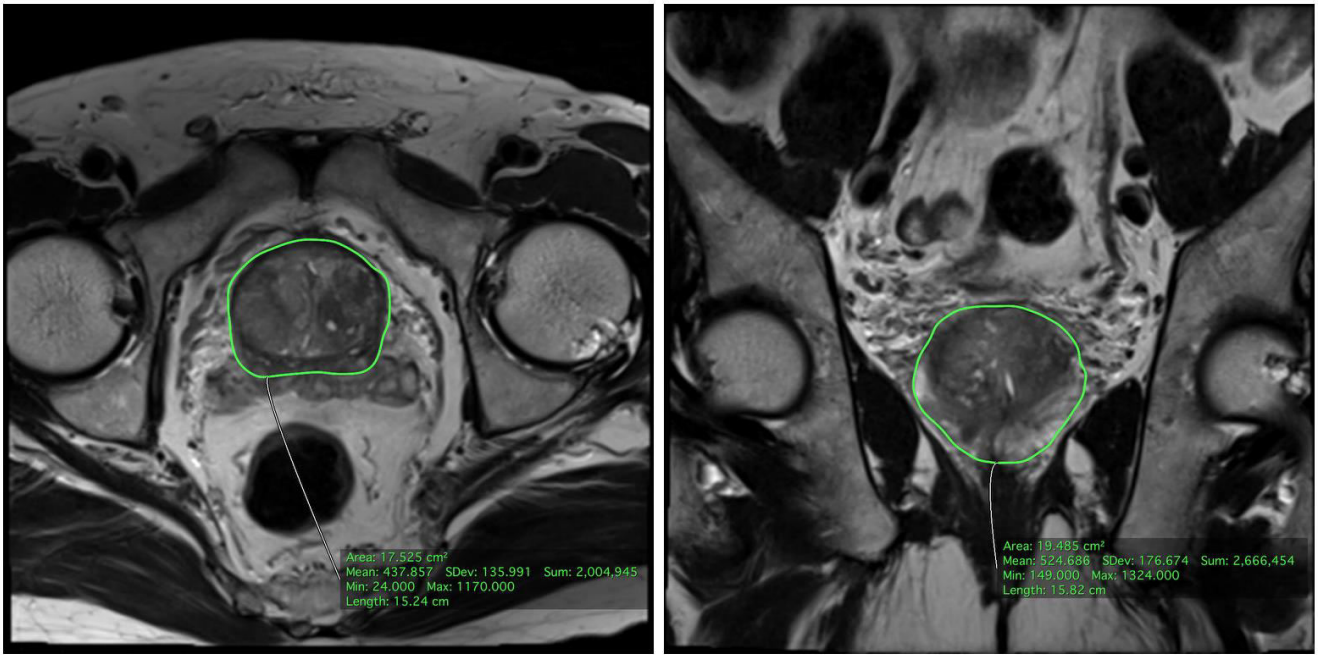

Figure-1. High resolution T2w MRI of PCa (a) Axial view; (b) Coronal view [21].

The VGG-16, VGG-19, and MobileNet-v3 models, all of which are pre-trained networks, were utilized in this research to classify the PCa. VGG is an abbreviation for Visual Geometry Group, and it is the deep CNN architecture with numerous layers that is standard in the industry. The term "deep" refers to the number of layers in a VGG-16 or VGG-19, which are respectively composed of 16 and 19 convolutional layers, respectively. Image analysis capabilities in many popular mobile applications are powered by MobileNet-v3, which is the third iteration of the architecture that underpins them. When it comes to classification tasks, MobileNet-v3 is both faster and more accurate than MobileNet-v2.

The rest of the work is presented in the following sections. Section two presents the related works for the research that covers the analysis of similar works. Section three presents the proposed research models, section four presents the results and analysis of the proposed models, and finally, section five presents the conclusion and future works.

2. RELATED WORKS

The 3D sliding window approach was implemented in [8] in relation to incorporating the 3D contextual spatial information provided by the MRI series while maintaining the 2D domain complexity and utilizing 3D data for PCa detection. This work proposed a deep convolutional encoder-decoder framework for segmenting and classifying the prostate as well as its malignant lesions and anatomical structure at the same time. It was not only conceivable, but also useful, to include 3D spatial information into a 2D multi-channel strategy using multiple channels. This fusion method enables the insertion of 3D context information into the 2D-based pipelines, which otherwise would not be possible. The encoder-decoder networks used for the segmentations of volumetric clinical images produce better outcomes than the previous method. Because of the restricted availability of large annotated datasets, it is not possible to compare and validate the results.

When it comes to PCa, the grading system can be thought of as ordinary class classification problems. The use of VGG-16 CNN and ordinary class classifier with J48 as the basis classifier in [9] was used to develop a method for grading PCa from mpMRI images. Using the kappa score of moderate quadratic weights, this approach could classify PCa belonging to distinct grade groups, indicating that it was effective. Some exact criteria were used in this

work to demonstrate that the Gleason grading does not allow for the low-aggressive malignancies recognition, and the cancer grading subjectivity results in important intra and inter-observing variability. The diagnosis accuracy achieved by this method needs to be improved even further in relation to being used efficiently in clinical settings.

The segmentation and classification of PCa were accomplished using a customized CNN architecture, which was created in [10]. With the ultimate goal of determining whether DTL, test-time augmentation, and unsupervised pre-training can improve pixel-by-pixel classifications, this model was combined with sophisticated approaches for handling classification issues (Non-cancer vs. Cancer). A considerable improvement in CNN performance was achieved by incorporating test-time augmentations with DTL or unsupervised pre-training. It would have been better if the size of the training and testing datasets had been increased in order to enhance the generalizability of the model.

In terms of automatic CAD analysis of PCa with mpMRI, deep learning appears to be the most promising methodology. In [11], the optimized patch-based DCNN system based on VGG-Net was constructed, and an improved prediction approach was included in the prediction phase in order to increase the prediction accuracy of the system. The number of images used for training was below 200, according to the researchers. More data would lead to more accurate results. When it came to diagnosing PCa from non-cancerous images, the model produced underfitting and overfitting results, respectively.

According to [12], the CNN and LADTree classifiers were proposed for evaluating PCa based on MRI data. This methodology was used to automate the gradings of PCa into five separate graded groups based on the Gleason grading groups, and it was successful. MRI images with T2 weighting, apparent diffusion coefficient (ADC), and high B-value diffusion weighting were employed in the evaluation. This model was unable to predict any cancers in the GG 4 grades. The inability to classify GG 4 lesions was due to an imbalance between classes and a lack of training data for the classification. The problem of class imbalance may have been resolved by employing any number of data augmentation approaches to increase the amount of training data available.

Prostate MRI is a technically challenging procedure that necessitates great image quality in order to achieve its full diagnostic potential. Image quality could be improved through the use of the automated approach for identifying diagnostically poor images, for example. A CNN-based analysis pipeline to classify the prostate MRI images quality was developed in [13]. This model exhibited great accuracy in prostate MRI classification of image quality on the individual-slice basis and nearly flawless accuracy in classifying whole sequences when applied to the entire dataset. The sample size for training in this study was rather small, which was a restriction that had the effectual performances of the model.

According to [14], a two-stage classification model for the identification of PCa has been presented. Extraction of characteristics for further diagnosis was accomplished through the use of trained models such as MobileNet and DenseNet. To make the forecast, the features gathered from these models were piled on top of one another. The stacked characteristics retrieved from the models were fed into the two-stage classification system, which aids in the classification process by making it more efficient. The logistic regression model was employed as the meta-classifier, and the features were fed into it in a pipelined way in order to acquire the corresponding outcomes. In this case, the logistic regression model serves as the main classifier, and the classification output was provided by the model.

Deep learning approaches have significantly enhanced the performance of prostate MRI analysis and classification in recent years. It was investigated in [15] if it was possible to utilize the Semantic Learning Machine (SLM) neuro-evolution algorithm for replacing the backpropagation method usually utilized in the last fully connected (FC) layers of CNN instead of the backpropagation algorithm. The multispectral MRI non-contrast-enhanced sequences from the data sets were evaluated in order to make a determination. It was critical to emphasize the significance of the data obtained by SLM without first training the CNN. This was a significant benefit over other current techniques and procedures because the SLM was not constrained or based on backpropagation, which is a significant advantage over other existing models and approaches.

In [16], a Textured-deep learning technique was proposed for the automatic PCa classification using 3T mpMRI, which was based on textured deep learning. It has been demonstrated that a texture-based CNN model for automatic PCa classifications of the prostate lesions was created using data from three-tesla (3T) MRI along with wholemount histopathology (WMHP) correlation. Following the detection and contouring of a lesion as the segment of the clinical interpretations, this deep learning model was created in order to additionally enhance the classifications of PCa for every positive MRI result in the future. This work was primarily concerned with demonstrating the benefits of combining GLCM-based texture data with CNN in the classifications of PCa.

Using T2W image-derived textural features, the potential for quantitative assessments of peripheral zones PCa aggressiveness was investigated in [17]. In this study, tumor volumes interpreted based on WMHP were used to construct the T2W intensity and ADC histograms, as well as T2W textural characteristics. Using Spearman correlations, this work is able to determine the relationship between textural characteristics and PCa grade sets. Textural aspects of T2W might have the potential to enhance PCa classifications in multicenter settings, according to recent research findings.

3. PROPOSED METHODOLOGY

In this proposed research, the PCa is classified based on the T2w-MRI data using DTL models. The

proposed models are based on DL architecture, which is used for the classification of PCa. From the dataset, a subset of MR Images of 845 patient records with a unique "UCLA" score of the ROI was used for multi-class classification. For classifying the data, the three different transfer learning models are used and these DTL models were trained using various optimizers for varying the learning rate. Finally, based on the results, the performance was compared among the different proposed models for validation.

A. VGG 16 and 19

Deep learning is widely regarded as having the ability to perform extremely accurate medical image classification, and this is widely acknowledged. In contrast to standard machine learning techniques, deep learning has several important advantages. The most notable advantage is that it eliminates the need for manual feature extraction because deep learning is capable of automatically extracting features and then classifying PCa. A key premise of DTL was for training a pre-trained CNN to utilize the smaller data set from a diverse problem to learn new image depictions rather than using the larger dataset from the same problem. DTL is a system for machine learning architecture; it was not a machine learning model or approach from a technical view. This design technique is typically used for models that have already been trained. Deep Evolution Neural Networks (DENNs) are used to train these pre-trained models. This technique, which is used in DL, includes the initial training of the CNN utilizing larger-scale training data sets for the purpose of solving a classification problem. In Figure 2, the architectures of the VGG-16 and 19 models have been presented.

As depicted in Figure 2, VGG is composed of six primary segments, the majority of which are multi-connected convolutional layers and FC layers combined. The dimension of the convolutional kernels is 3*3. Generally, every layer was concentrated at 16~19, with 16 being used for the VGG-16 and 19 being used for the VGG-19 models. The model is composed of largely connected FC and convolutional layers, which allows for improved feature extractions as well as the utilization of Maxpooling (instead of averaged pooling) for downsampling prior classifications using the SoftMax activation functions, which improves the overall classification performance [18].

In the VGG-16 algorithm, the input image is routed through an array of convolutional layers each with a receptive field of three. The stride of the convolution is taken to be one pixel. Spatial pooling (downsampling) is accomplished by the use of five max-pooling layers with strides equal to two. Max-pooling layers are applied after a few of the convolution layers, and the operation was carried out over the 22-pixel window. Following the sets of conv layers, there were three FC layers, each of which has channel sizes of 4096, 4096, and 1000 bits. During the FC layers, every neuron receives information from the activation of the neuron in the layers above it.

The expression for the downsampling layer is given in the following equation (1),

$$X_{p_j}^{(n)} = f\left(\tau_j^n down\left(X_j^{(n-1)}\right) + b_j^{(n)}\right) \quad (1)$$

According to equation (2), the activation function of ReLU is represented as,

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \quad (2)$$

It is possible to express the activation function of the softmax layer as follows in equation (3),

$$f\left(x_j\right) = \frac{e^{x_j}}{\sum e^{x_i}} \quad (3)$$

In equation (3), $f(x)$ denotes the activation functions, and x denotes the input to the activation functions.

The VGG-19 network has 19 weight layers, 16 of which were convolution layers and three of which are forward convolutional layers. All of the convolutional layers make use of three-way convolution kernels. To imitate a bigger receptive field effect, convolution kernels are continually stacked on top of one another to generate a convolution sequence every 2 or 4 convolution kernels. When the 22pooling window was used in the pooling layers, the translation invariance of the model can be maintained while also allowing the sizes of the feature maps after convolutions to be reduced, which can be beneficial. Lastly, the FC layer is combined with three continuous FC layers, with the number of channels being 4096, 4096, and 1000, before being classed and output by the Softmax classifier with a total of 1000 labels [19].

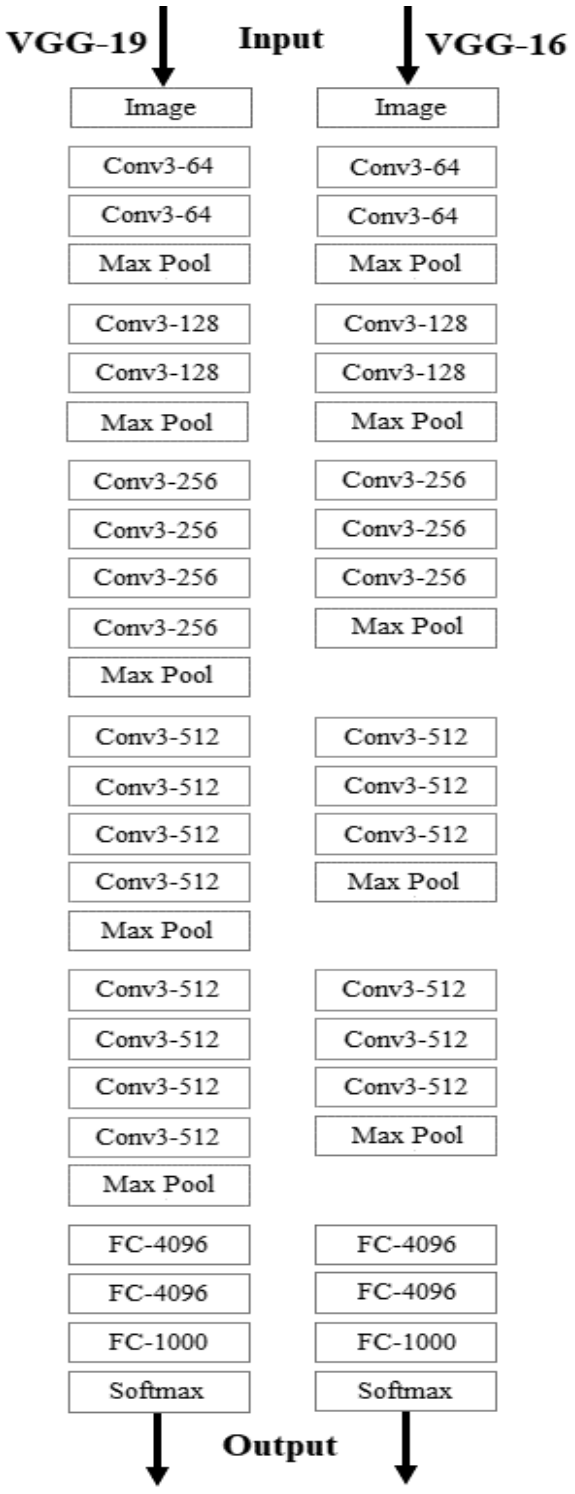

Figure-2. Architecture of VGG-16 and VGG-19.

B. MobileNet-v3

With the integration of hardware-aware network architecture search (NAS) and the NetAdapt algorithm, MobileNet-v3 is a CNN that was tuned to mobile phone CPUs and then later improved through innovative architecture advancements. A proposal for MobileNet-v3 was made by the Google team in the year 2019. NAS-developed mobile neural architecture search network (MnasNet) serves as the foundation for its network design. MobileNet-v3 is described as two models: MobileNet-v3-Large and MobileNet-v3-Small. MobileNet-v3-Large is the larger of the two models. These models are intended for scenarios with a high demand for resources and cases with a low demand for resources. The models are constructed by utilizing platform-aware NAS and NetAdapt for network search, as well as by incorporating network improvements. In this work, the MobileNet-v3Large model is used.

Input	Operator	exp size	#*out*	SE	NL	*s*
$224^2 \times 3$	conv2d	-	16	-	HS	2
$112^2 \times 16$	bneck, 3x3	16	16	-	RE	1
$112^2 \times 16$	bneck, 3x3	64	24	-	RE	2
$56^2 \times 24$	bneck, 3x3	72	24	-	RE	1
$56^2 \times 24$	bneck, 5x5	72	40	✓	RE	2
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 5x5	120	40	✓	RE	1
$28^2 \times 40$	bneck, 3x3	240	80	-	HS	2
$14^2 \times 80$	bneck, 3x3	200	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3x3	480	112	✓	HS	1
$14^2 \times 112$	bneck, 3x3	672	112	✓	HS	1
$14^2 \times 112$	bneck, 5x5	672	160	✓	HS	2
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	bneck, 5x5	960	160	✓	HS	1
$7^2 \times 160$	conv2d, 1x1	-	960	-	HS	1
$7^2 \times 960$	pool, 7x7	-	-	-	-	1
$1^2 \times 960$	conv2d 1x1, NBN	-	1280	-	HS	1
$1^2 \times 1280$	conv2d 1x1, NBN	-	k	-	-	1

Figure-3. Architecture specification of MobileNet-v3-large [20].

As mentioned in Figure-3, the presence or absence of the Squeeze-and-Excitation in a block is indicated by the term *SE*. The *NL* represents the nonlinearity type that was employed. The terms *HS* and *RE* stand for h-swish and ReLU, respectively. The abbreviation *NBN* refers to no batch normalization. The term *s* stands for stride.

The deep separable convolution layer from MobileNet-v1 as well as the inverse residual structure with a linear bottleneck layer from MobileNet-v2 were added to this MobileNet-v3 model. A nonlinearity known as swish has been introduced, and when employed as the drop-in replacement for ReLU, it has been shown to dramatically increase the neural network model's accuracy. The nonlinearity is defined in the following equation (4).

$$swish\ x = x \cdot \sigma(x) \tag{4}$$

However, while this NL enhances accuracy, it comes at a non-zero cost in embedded systems because the sigmoid function was far more costly to calculate than the sigmoid function used in the simulation. Additionally, to lower the hardware resource utilization in the embedded context, a new activation function *h-swish*(x) was developed to estimate the replacements of the sigmoid, as seen in the equation below.

$$h - swish\ (x) = x\frac{ReLU6(x+3)}{6} \quad (5)$$

Prior to the average pooling layers, there was a 1x1 convolutional layer, which could enhance the dimensions of the feature maps and make predictions easier, but it further adds to the computing burden by increasing the number of convolutional layers. As a result, MobileNet-v3 places it behind average pooling in terms of performance. First, the average pooling layer was utilized for reducing the features map sizes from 7x7 to 1x1, and hence to 1x1, and so on, drastically reducing the amount of computing required. Convolutions of 3x3 and 1x1 are directly extracted from the previous convolution to significantly minimize the total computing required to do it. There is a significant increase in network operational efficiency without a significant decrease in accuracy [20].

MobileNet-v3 was partitioned into three sections, the first of which comprises one convolutional layer, followed by three layers of 3x3 convolutions to extricate features. The second part comprises two convolution layers, and the third part includes three layers of 3x3 convolution to extract features. In the following section, many bottle-neck convolutional layers are included, which are formed of multiple convolutions, deep wise convolutions, SE structures, and other elements, and in the final section, two 1x1 convolutional layers are used to create class output.

All these proposed models of VGG-16, VGG-19, and MobileNet-v3 were trained using different optimizers like Adam, AdaMax, SGD, RMSprop, and Ftrl techniques. Based on these five optimizers five models were developed on VGG-16, VGG-19, and MobileNet-v3.

- Model-01 was modified with Adam optimizer with a 0.001 learning rate for both VGG-16 and 19 and 0.00001 learning rate for MobileNet-v3.
- Model-02 was modified with an SDG optimizer with a 0.001 learning rate for both VGG-16 and 19 and a 0.00001 learning rate for MobileNet-v3.
- Model-03 was modified with RMSprop optimizer with a 0.00001 learning rate for both VGG-16, VGG-19, and MobileNet-v3.
- Model-04 was modified with an Ftrl optimizer with a 0.001 learning rate for both VGG-16 and 19 and a 0.00001 learning rate for MobileNet-v3.
- Model-05 was modified with AdaMax optimizer with a 0.001 learning rate for both VGG-16 and 19 and a 0.00001 learning rate for MobileNet-v3.

4. RESULTS AND DISCUSSIONS

The proposed model was implemented and tested using the PYTHON 3.7.12 programming language tool. The experiments are performed on Google Colab Pro. The dataset was collected from the TCIA website given by the University of California, Los Angeles (UCLA), and used in this study for evaluation purposes.

A. Dataset Description

In this Prostate-MRI-US-Biopsy dataset, US and MR images of 1151 individuals as well as 61,119 DICOM images, for a total image dataset size of 77.6 GB (GB). The mpMRI sequences, such as T2W, DW, and DCE, were used to establish MRI targets, which were then rated on a Likert-like scale that was closely related to the PIRADS version 2 scale. The ROI outlines were traced using t2-weighted MRI, which was the only sequence included in this data set. The MRI was carried out on a Verio, or Skyra and 3 Tesla Trio scanners, depending on the manufacturer (Siemens, Erlangen, Germany). All cases were treated with a transabdominal phased array, with an endorectal coil being employed in a subset of instances. 3D T2: SPC pulse sequences account for the vast majority of pulse sequences, with TR/TE of 2200/203, Matrix/FOV of 256 205/14 14 cm, and 1.5mm slice spacing being the most common [22]. The dataset can be downloaded from this link Prostate MRI and Ultrasound with Pathology and Coordinates of Tracked Biopsy (Prostate-MRI-US-Biopsy).

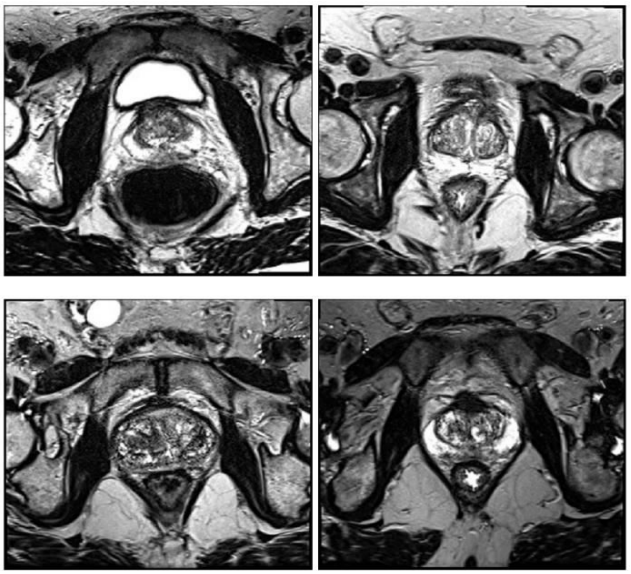

Figure-4. Sample images from dataset.

This PCa dataset is made available to the public through a data-sharing platform. In order to perform Multi-Class Classification, a subset of T2W MR images from 845 patient records that had a unique "UCLA" score of the ROI was selected from the dataset and used for evaluation. The data set was split into two parts: 70 percent for training and 30 percent for testing, which was used for performance analysis.

The Likert-like score was a visual analog scale on which scores of 1-5 signify the chance of cancer in the prostate based on the total interpretations of the prostate MRI by a radiologist, as opposed to a numerical scale.

Scores 1 and 2 indicate a low suspicion of cancer, a score of 3 indicates an equivocal suspicion, and a score of 4 or 5 indicates a high suspicion of cancer. In this case, zero indicates the normal/healthy image.

B. Performance Metrics

For evaluating the performances of the proposed models and experimental outcomes, this work included performance measures like accuracy, sensitivity, precision, specificity, and F1 scores. The terms TP (True Positive) and TN (True Negative) indicate the proper identification of PCa, while the terms FP (False Positive) and FN (False Negative) indicate the incorrect identification of PCa.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (8)$$

$$Specificity = \frac{TN}{TN+FP} \quad (9)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (10)$$

The proposed models of VGG-16, VGG-19, and MobileNet-v3 were trained on multiple optimizers with different learning rates. The accuracy, precision, sensitivity, specificity, and F1-scores were calculated based on the evaluation of the models performed on the prostate dataset images.

Table-1. Performance evaluation of VGG-16 Model-04.

Class	Accuracy	Precision	Sensitivity	Specificity	F1-Score
0	0.9331	0.75	1	0.9993	0.8571
1	0.9331	0.8593	0.9649	0.9944	0.9090
2	0.9331	0.9444	1	0.9993	0.9714
3	0.9331	0.9796	0.9150	0.9842	0.9462
4	0.9331	0.9128	0.9490	0.9601	0.9306
5	0.9331	0.8888	0.9399	0.9665	0.9137

The performance analysis of classifying the PCa using VGG-16 with multiple optimizers was tabulated. By using the five different optimizers, five different VGG-16 models were developed for the classification of PCa. But in this paper, out of five, two of the best-performed model's evaluations VGG-16_Model-04 and VGG-16_Model-05 only presented as shown in Tables 1 and 2. Model-04 was modified with Ftrl optimizer with 0.001 learning rate and Model-05 was modified with AdaMax optimizer with 0.001 learning rate. Other models like Model-01, 02, and 03 performed below these presented models.

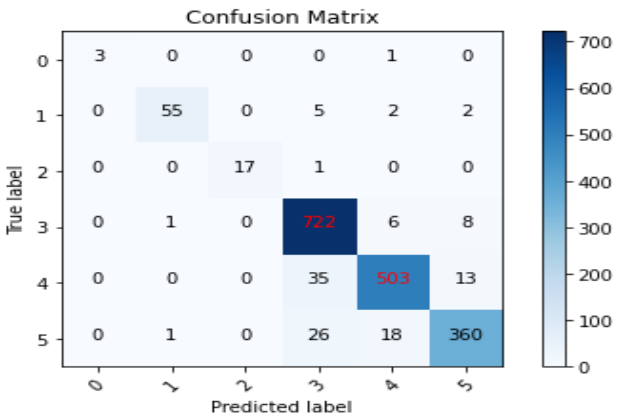

Figure-5. Confusion matrix of VGG-16 Model-04.

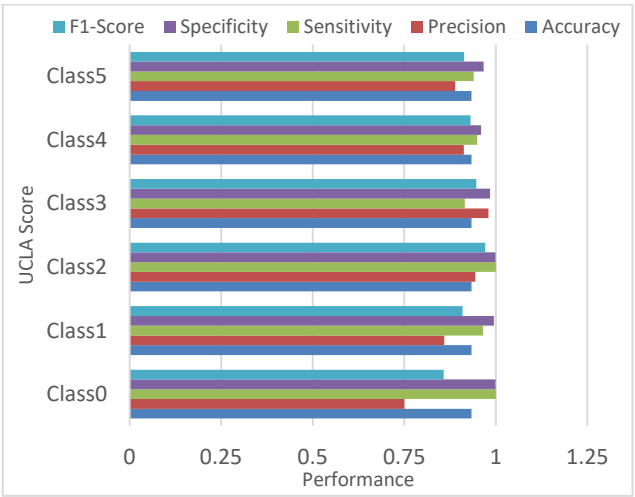

Figure-6. Performance plot of VGG-16 Model-04.

Comparing models VGG-16_Model-04 and VGG-16_Model-05, Model-04 obtained the best performance with a constant accuracy of 93.31% in classifying PCa from the given dataset images. Figures 5 and 7 represent the confusion matrix of these VGG-16 models and the Figures 6 and 8 represent the graphical plots of these VGG-16 models based on the performance analysis.

Table-2. Performance evaluation of VGG-16 Model-05.

Class	Accuracy	Precision	Sensitivity	Specificity	F1-Score
0	0.8954	0.75	1	0.9993	0.8571
1	0.8954	0.7187	0.92	0.9884	0.8070
2	0.8954	0.8333	1	0.9981	0.9090
3	0.8954	0.9308	0.9038	0.9467	0.9171
4	0.8954	0.8983	0.8855	0.9514	0.8918
5	0.8954	0.8592	0.8854	0.9562	0.8721

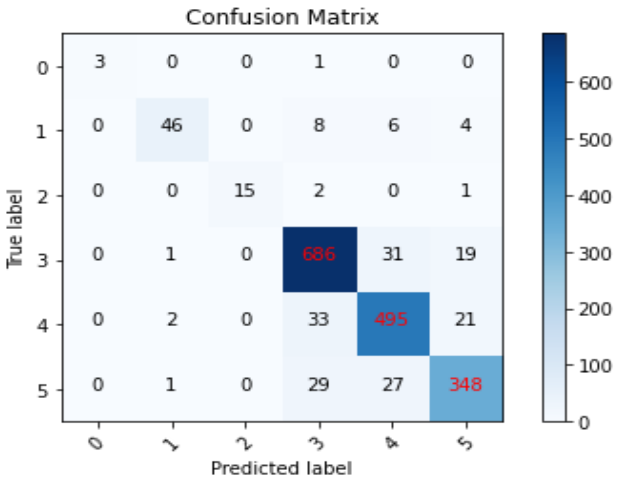

Figure-7. Confusion matrix of VGG-16 Model-05

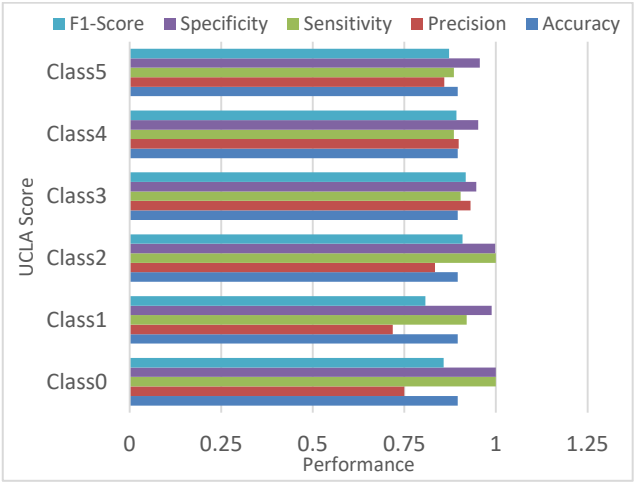

Figure-8. Performance plot of VGG-16 Model-05.

Table-3. Performance evaluation of VGG-19 Model-04

Class	Accuracy	Precision	Sensitivity	Specificity	F1-Score
0	0.9392	1	1	1	1
1	0.9392	0.8125	0.9629	0.9926	0.8813
2	0.9392	0.8888	1	0.9987	0.9411
3	0.9392	0.9538	0.9474	0.9660	0.9506
4	0.9392	0.9491	0.9306	0.9761	0.9398
5	0.9392	0.9209	0.9301	0.9759	0.9255

In Tables 3 and 4, the performance analysis of classifying the PCa using VGG-19 with ftrl and AdaMax optimizers were tabulated. Out of five models, two of the best-performed model's evaluations VGG-19_Model-04 modified with Ftrl optimizer with 0.001 learning rate and VGG-19_Model-05 modified with AdaMax optimizer with 0.001 learning rate only presented. Other models like Model-01, 02, and 03 performed below these presented models.

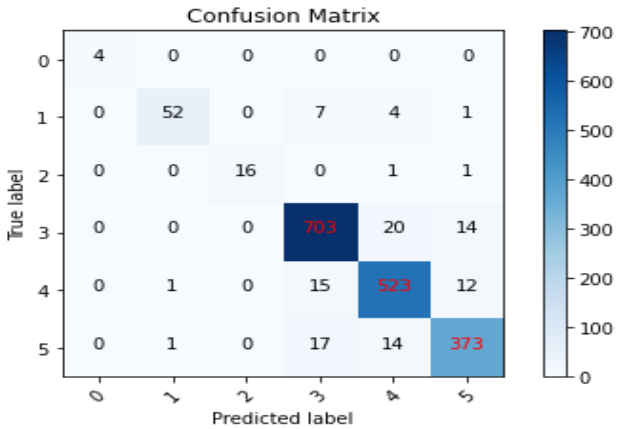

Figure-9. Confusion matrix of VGG-19 Model-04.

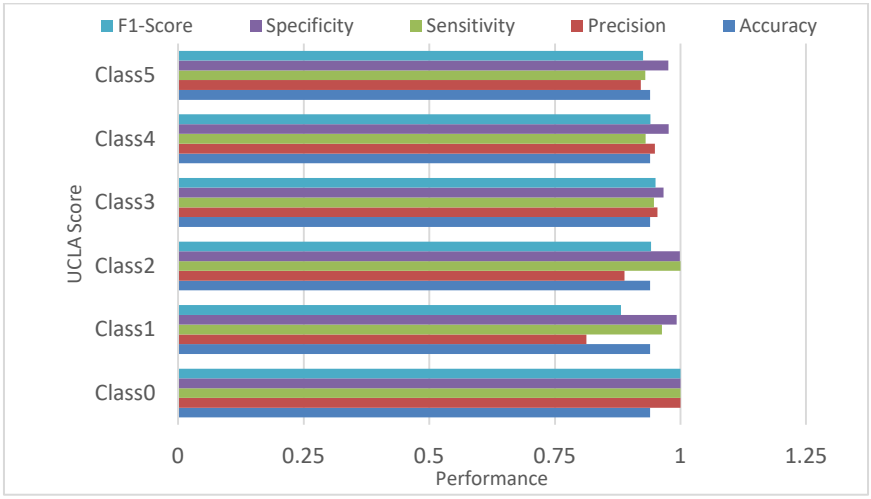

Figure-10. Performance plot of VGG-19 Model-04.

By comparing the models VGG-16_Model-04 and VGG-16_Model-05, Model-04 obtained the best performance with a constant accuracy of 93.92% in classifying PCa from the given dataset images. Figures 9 and 11 represent the confusion matrix of these VGG-19 models and figures 10 and 12 represents the graphical plots of these VGG-19 models based on the performance analysis. Compared to the best-performed VGG-16 model-04, the VGG-19 model-04 obtained better performance and achieved better accuracy.

Table-4. Performance evaluation of VGG-19 Model-05.

Class	Accuracy	Precision	Sensitivity	Specificity	F1-Score
0	0.8572	1	0.8	1	0.8888
1	0.8572	0.7812	0.8333	0.9905	0.8064
2	0.8572	1	0.9	1	0.9473
3	0.8572	0.8887	0.8863	0.9138	0.8875
4	0.8572	0.8294	0.8638	0.9191	0.8462
5	0.8572	0.8419	0.8004	0.9487	0.8206

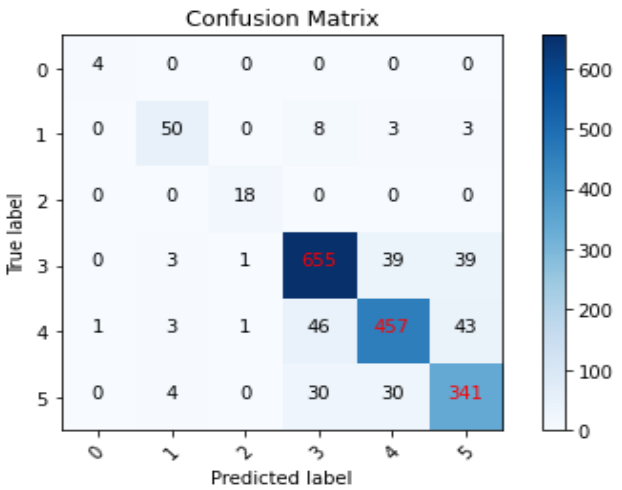

Figure-11. Confusion matrix of VGG-19 Model-05.

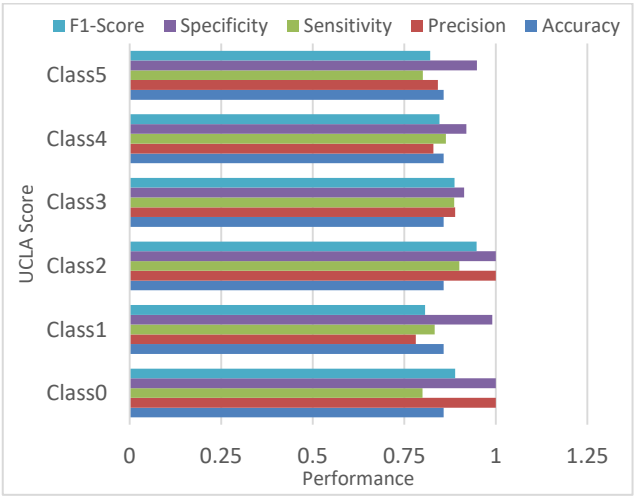

Figure-12. Performance plot of VGG-19 Model-05.

Table-5. Performance evaluation of MobileNet-v3 Model-01.

Class	Accuracy	Precision	Sensitivity	Specificity	F1-Score
0	0.9213	1	1	1	1
1	0.9213	0.7343	1	0.9894	0.8468
2	0.9213	0.8333	1	0.9981	0.9090
3	0.9213	0.9131	0.9438	0.9378	0.9282
4	0.9213	0.9564	0.8667	0.9788	0.9094
5	0.9213	0.9209	0.9515	0.9753	0.9360

As shown in Tables 5 and 6, the performance analysis of classifying the PCa using MobileNet-v3 with Adam and Ftrl optimizers was tabulated. Out of five models, two of the best-performed model's evaluations MobileNet-v3_Model-01 modified with Adam optimizer with 0.00001 learning rate and MobileNet-v3_Model-04 modified with Ftrl optimizer with 0.00001 learning rate only presented. Other models like Model-02, 03, and 05 performed below these presented models.

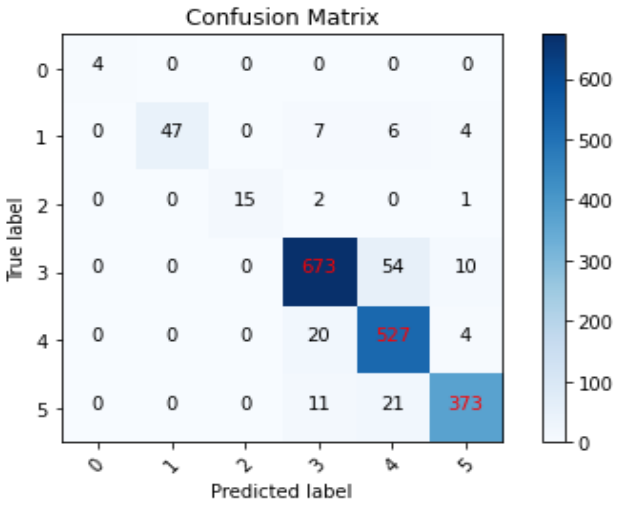

Figure-13. Confusion matrix of MobileNet-v3 Model-01.

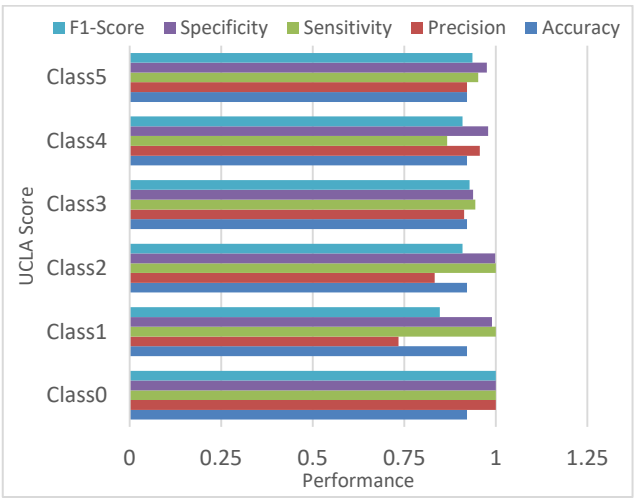

Figure-14. Performance plot of MobileNet-v3 Model-01.

When comparing the models MobileNet-v3_Model-01 and MobileNet-v3_Model-04, Model-04 obtained the best performance with a constant accuracy of 95.27% in classifying PCa from the given dataset images. Figures 13 and 15 represent the confusion matrix of these MobileNet-v3 models and figures 14 and 16 represent the graphical plots of these MobileNet-v3 models based on the performance analysis. Compared to the best-performed VGG-16 model-04 and VGG-19 model-04, the proposed MobileNet-v3_Model-04 obtained better performance and achieved better accuracy.

Table-6. Performance evaluation of MobileNet-v3 Model-04.

Class	Accuracy	Precision	Sensitivity	Specificity	F1-Score
0	0.9527	1	1	1	1
1	0.9527	0.8437	0.9818	0.9939	0.9075
2	0.9527	0.9444	1	0.9994	0.9714
3	0.9527	0.9660	0.9582	0.9751	0.9621
4	0.9527	0.9673	0.9334	0.9847	0.9500
5	0.9527	0.9259	0.9640	0.9777	0.9445

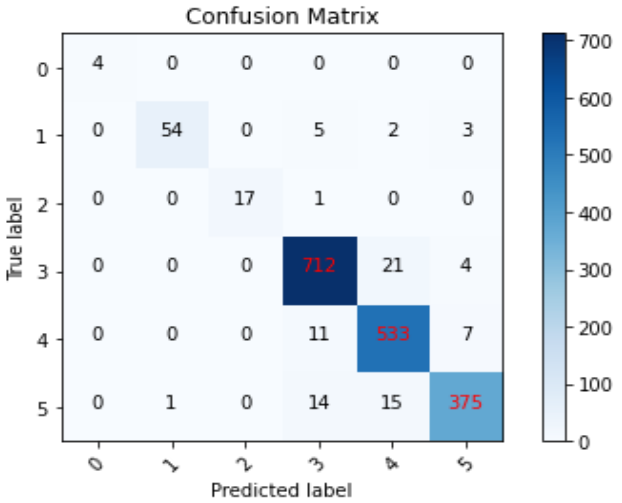

Figure-15. Confusion matrix of MobileNet-v3 Model-04.

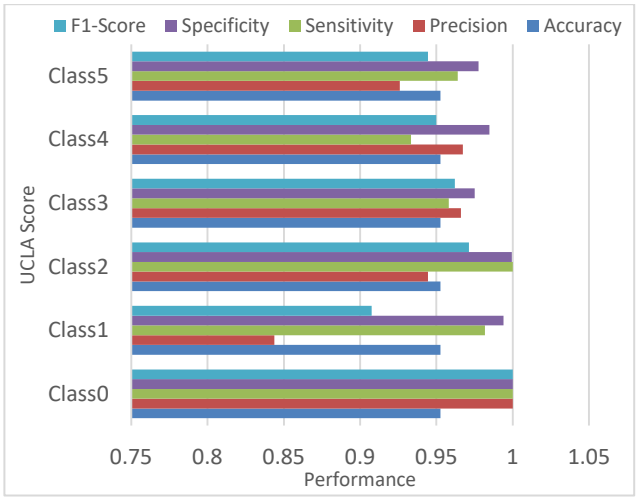

Figure-16. Performance plot of MobileNet-v3 Model-04.

5. CONCLUSIONS

In this research, DTL models such as VGG-16, VGG-19, and MobileNet-v3Large architectures were proposed to classify the PCa with the help of T2w MRI dataset images. These proposed models were modified using different optimizers with different learning rates. Optimizers like Adam, AdaMax, SGD, RMSprop, and Ftrl are used in this research. These models were fine-tuned by most of the top convolution layers and FC layers. The Prostate-MRI-US-Biopsy dataset was used for the evaluation of the proposed models, which contains both US and MR Images of 1151 patients, and 61,119 DICOM Images. From the dataset, a subset of T2w MR Images of 845 patient records with a unique "UCLA" score of the ROI was used for Multi-Class Classification. For performance analysis, the data set was split into 70% for training and 30% for testing. Although a series of different optimizers were used to optimize the proposed models with different learning rates, the models optimized with Ftrl optimizer performed well obtaining the best performances compared to the other proposed models. According to the obtained performances, the proposed models optimized using Ftrl optimizer have obtained better performances with 93.31% accuracy, 93.92% accuracy, and 95.27% accuracy for VGG-16-Model-04, VGG-19-Model-04, and MobileNet-v3 respectively. The MobileNet-v3 modified with Ftrl optimizer obtained the best performance with higher accuracy than other models. The last performance was obtained from the VGG-16. In the future, the proposed research can be focused on validating the performances of the models with diverse datasets and other existing methods.

REFERENCES

[1] Lizhi Liu *et al*. 2016. Computer-aided detection of prostate cancer with MRI: technology and applications. Academic Radiology. 23(8): 1024-1046, 2016.

[2] Hyuna Sung *et al*. 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians 71(3): 209-249.

[3] Renato Cuocolo i. 2019. Machine learning applications in prostate cancer magnetic resonance imaging. European radiology experimental, 3(1): 1-8, 2019.

[4] Jose M. Castillo T *et al*. 2022. Classification of clinically significant prostate cancer on multi-parametric MRI: A validation study comparing deep learning and radiomics. Cancers. 14(12): 1-15.

[5] Samira Masoudi *et al*. 2021. Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. Journal of Medical Imaging, 8(1): 1-14, 2021.

[6] Gabriel Nketiah *et al*. 2017. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: preliminary results. European radiology. 27(7): 3050-3059.

[7] Narmatha C. and Prasad M. Surendra. A Review on Prostate Cancer Detection using Deep Learning Techniques. Journal of Computational Science and Intelligent Technologies. 1(2): 26-33.

[8] RubaAlkadi *et al*. 2019. A deep learning-based approach for the detection and localization of prostate cancer in T2 magnetic resonance images. Journal of digital imaging. 32(5): 793-807, 2019.

[9] Bejoy Abraham and Madhu S. Nair. 2019. Automated grading of prostate cancer using convolutional neural

network and ordinal class classifier. Informatics in Medicine Unlocked. 17, 100256.

[10] David Hoar *et al.* 2021. Combined Transfer Learning and Test-Time Augmentation Improves Convolutional Neural Network-Based Semantic Segmentation of Prostate Cancer from Multi-Parametric MR Images. Computer Methods and Programs in Biomedicine, 210, 106375, 2021.

[11] Yang Song *et al*. 2018. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. Journal of Magnetic Resonance Imaging. 48(6): 1570-1577.

[12] Bejoy Abraham and Madhu S. Nair. 2019. Computer-aided grading of prostate cancer from MRI images using convolutional neural networks. Journal of Intelligent & Fuzzy Systems. 36(3):2015-2024.

[13] Stefano Cipollari *et al.* 2022. Convolutional neural networks for automated classification of prostate multiparametric magnetic resonance imaging based on image quality. Journal of Magnetic Resonance Imaging. 55(2):480-490.

[14] Jisha John, Aswathy Ravikumar and Bejoy Abraham 2021. Prostate cancer prediction from multiple pretrained computer vision model. Health and Technology. 11(5): 1003-1011.

[15] Paulo Lapa *et al*. 2019. Semantic learning machine improves the CNN-based detection of prostate cancer in non-contrast-enhanced MRI. Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1837-1845, 2019.

[16] Yongkai Liu *et al*. 2021. Textured-Based Deep Learning in Prostate Cancer Classification with 3T Multiparametric MRI: Comparison with PI-RADS-Based Classification. Diagnostics. 11, 1785.

[17] Gabriel A. Nketiah *et al*. 2021. Utility of T2-weighted MRI texture analysis in assessment of peripheral zone prostate cancer aggressiveness: A single-arm, multicenter study. Scientific Reports. 11, 2085.

[18] Ikechukwu A. Victor *et al.* 2021. ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images. Global Transitions Proceedings. 2(2): 375-381.

[19] Manimurugan S. 2020. Classification of Alzheimer's disease from MRI Images using CNN based Pre-trained VGG-19 Model. Journal of Computational Science and Intelligent Technologies. 1(2): 34-41.

[20] Andrew Howard *et al.* 2019. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1314-1324.

[21] Zia Khan *et al*. 2021. Recent Automatic Segmentation Algorithms of MRI Prostate Regions: A Review. IEEE Access. 9: 97878-97905.

[22] Natarajan S. *et al*. 2020. Prostate MRI and Ultrasound with Pathology and Coordinates of Tracked Biopsy (Prostate-MRI-US-Biopsy) Dataset.