IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 20, 2021, accepted January 29, 2021, date of publication February 8, 2021, date of current version February 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3057654

Prostate Cancer Detection Using Deep Learning and Traditional Techniques

SAQIB IQBAL[1], GHAZANFAR FAROOQ SIDDIQUI[1], AMJAD REHMAN[2], (Senior Member, IEEE), LAL HUSSAIN[3,4], TANZILA SABA[2], (Senior Member, IEEE), USMAN TARIQ[5], AND ADEEL AHMED ABBASI[3]

[1]Department of Computer Science, Quaid-i-Azam University, Islamabad 15320, Pakistan
[2]Artificial Intelligence and Data Analytics Lab, CCIS, Prince Sultan University, Riyadh 11586, Saudi Arabia
[3]Department of Computer Science and IT, University of Azad Jammu and Kashmir, City Campus, Muzaffarabad 13100, Pakistan
[4]Department of Computer Science and IT, University of Azad Jammu and Kashmir, Neelum Campus, Athmuqam 13100, Pakistan
[5]College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Alkharj 11942, Saudi Arabia

Corresponding authors: Ghazanfar Farooq Siddiqui (ghazanfar@qau.edu.pk) and Usman Tariq (u.tariq@psau.edu.sa)

ABSTRACT Prostate cancer (PCa) is a severe type of cancer and causes major deaths among men due to its poor diagnostic system. The images obtained from patients with carcinoma consist of complex and necessary features that cannot be extracted readily by traditional diagnostic techniques. This research employed deep learning long short-term memory (*LSTM*) and Residual Net (*ResNet* − 101), independent of hand-crafted features, and is fine-tuned. The results were compared with hand-crafted features such as texture, morphology, and gray level co-occurrence matrix (*GLCM*) using non-deep learning classifiers such as support vector machine (*SVM*) Gaussian Kernel, k-nearest neighbor-Cosine (*KNN* − *Cosine*), kernel naive Bayes, decision tree (*DT*) and RUSBoost tree. This study reduces the features of carcinoma images, employed machine learning and deep learning approaches. For validation of training and testing data, a jack-knife ten-fold cross-validation method was used. The performance was measured using a confusion matrix such as sensitivity, specificity, positive predictive value (*PPV*), negative predictive value (*NPV*), accuracy (*AC*), Mathews Correlation Coefficient (*MCC*), and area under the curve (*AUC*). The most remarkable performance was obtained using non-deep learning methods with *GLCM* features using KNN-Cosine with sensitivity (98.00%), specificity (99.25%), PPV (98.99%), NPV (99.11%), accuracy (99.07%), and AUC (0.998). The LSTM deep learning method yields performance with sensitivity (98.33%), specificity (100%), PPV (100%), NPV (99.26%), accuracy (99.48%), MCC (0.9879) and AUC (0.9999), where using Deep learning method *ResNet* − 101, we obtained (100%) Accuracy and AUC (1) for Kernel Naive Bayes, SVM Gaussian and RUSBoost Tree. The results show that *ResNet* − 101 deep learning outperformed than non-deep learning methods and *LSTM*. Thus, the deep learning method *ResNet* − 101 could be used as a better predictor for the detection of prostate cancer.

INDEX TERMS Deep learning, feature extraction, healthcare, *LSTM*, machine learning, morphological, prostate cancer, texture.

I. INTRODUCTION

The prostate is a walnut-sized male reproductive gland that produces and secretes fluid of an alkaline nature. The gland is located in the pelvis where the rectum is on the posterior side, and the bladder is on the upper side, the gland is surrounded by parts of the urethra [1]. The skeletal muscles inside extend from the diaphragm to the apex. It has different regions, the region near the bladder is named the base, and the part of the gland closest to the ureteral sphincter is called the apex [2]. Anatomically four lobs of the prostate are the peripheral zone (*PZ*), the transition zone (*TZ*), the central zone (*CZ*), and the anterior fibromuscular trauma, among the first three consists of 70%, 5%, and 25% tissue respectively. The latter consists of no glandular tissue [3]. Research shows that the higher the glandular tissue count, the higher the risk of prostate cancer [4]. The proliferation of cells causes prostate cancer.

The associate editor coordinating the review of this manuscript and approving it for publication was Shen Yin.

Cancer is a global health problem where patients suffer from clinical and management issues of cancer. Older people with less physical activity have a high risk of cancer. In the United States, 1, 806, 590 (nearly 1.8 million) new cancer cases and 606, 520 (0.6 million) deaths have been recorded in the current years, as reported by cancer statistics 2020 [5]. Statistics show that the death rate in four primary cancers (prostate, breast, lung, and colorectal) increased until 1990 and a gradual decrease until 2017. Prostate cancer is the most common type and the second leading death figure in the United States among men [6]. New cases of prostate cancer in the United States in 2019 were 191, 930 out of which 21% are men. Out of 893, 660 cases, the leading cancer type was prostate cancer [7]. The estimated deaths in 2020 due to cancer in the United States were 321, 160, of which 33, 310 were due to prostate cancer. Prostate cancer depends on age, family history, and lifestyle. Moreover, Asians are less vulnerable than Europeans and blacks. Active and reliable treatments such as radical prostatectomy, radiotherapy, and hormone therapy have reduced the mortality rate. As active radiation is an effective treatment procedure, it can cause psychological and emotional side effects [8].

There are several ways to diagnose prostate cancer; using screening tests for urinary symptoms, most men complain of chronic inflammation about these tests. However, many men go their entire lives without being diagnosed with prostate cancer [9]. The digital rectal exam (DRE) was used in the early 1990s to screen prostate cancer. A remarkable reduction has been reported due to its implication. This screening test is also used to distinguish between benign and malignant cells [10]. The DRE is useful for diagnosing tumors in the posterior area and cannot diagnose tumors in other parts or regions of the prostate [11]. Prostate-specific antigen (PSA) is used as a screening test. It was introduced in the early 1990 and can help reduce the mortality rate and other complications. Now it is proving to be very controversial. PSA screening often generates false-positive results and certain health complications, such as pain and infection [12]. PSA can be useful for asymptomatic men, where its consequences may go unrecognized for life. The systematic randomized method under transrectal ultrasound (TRUS) helps detect small and low-grade cancers used in clinical diagnosis. Still, its low sensitivity makes it difficult to screen a large population [13]. There are several methods to diagnose the prostate using computer-aided design (CAD). Their accuracy lies between 0.80 and 0.89, the authors obtained the AUC value of 0.95 to 0.96 in [9]. However, their system was based on the manual selection of candidates and data-dependent. The Ultra Sono-Grapy (USG) imaging modality is an imaging technology used in the initial detection based on real-time images. Still, this method cannot generate optimal results due to the limitation of the contrast between cancerous and benign tissue [14]. Multiparametric magnetic resonance imaging (mpMRI) is more robust than ultrasound because of its excellent tissue contrast, but due to the lack of real-time imagining, it requires advanced training [15].

Brachytherapy is a type of radiation therapy used to treat neck, skin, breast, and prostate cancer. In this approach, the radioactive beam is directly imposed on the affected tissue. Brachytherapy has two modes, permanent placement (also called seed placement) and removable radioactive material. Brachytherapy or the implementation of seeds is more effective in prostate cancer, replacing the radial prostatectomy. During operating the prostate, using Brachytherapy, pallets(small seeds) are placed on the prostate cell. These pallets are very tiny but radioactive. This highly radioactive material kills prostate cancer cells. As a side effect, it may also harm or even kill the healthy cells near the prostate.

Researchers have recently used deep and non-deep learning methods to detect anomalies in medical images. For non-deep learning, extracting hand-crafted features is the most critical step. To study the underlying system, different features have been extracted from a biological, physiological, and neurophysiological system in the past.

The non-deep learning methods are fine-tuned and independent of the hand-crafted features. We used LSTM and ResNet-101 from deep learning methods. Recently, LSTM blocks as a feature of RNNs are used for speech recognition, speech and language modeling, and cognitive performance classification by Greff *et al.* [16]. Magnetic resonance imaging (mMRI) for prediction and detection purposes using machine learning and deep learning. Moreover, Hussain *et al.* in [17] employed a novel method for the detection of prostate carcinoma, different scale-invariant feature transform (SIFT), morphological, elliptic Fourier descriptors (EFDs), texture, and entropy were computed from MRI images, using SVM, DT, and Bayes Classification.

The researchers of [18], [19] extracted hybrid features based on geometry, morphology, texture properties, noise-robust, elliptical shapes, and complex dynamics. They used robust machine learning techniques to predict different cancer types, such as colon cancer, lung cancer, prostate cancer, breast cancer, and brain tumor. A pipeline based on the convolutional neural network (CNN) has been proposed by [20] to detect prostate cancer; they achieved an AUC of 0.87. The authors of [21] also presented a deep learning-based architecture (x masNet) to classify prostate cancer. The results show that the deep learning approach has great potential for detecting prostate cancer from medical images. Authors in [22] use $ResNet - 101$ for segmentation of coronavirus ($COVID - 19$) effected regions of medical images. They obtain 0.85 $F1$ score. Authors in [23] classify brain tumer using $ResNet - 101$ deep method. Moreover, the authors of [24] found that machine learning techniques are more optimal and robust for detecting prostate cancer than the antigen density and velocity. Likewise, [25] has proposed a deep learning model to detect prostate lesions in magnetic resource images (MRI) by achieving an AUC of 0.955, which may improve the cancer diagnosis. Machine learning methods are based on handed features, where deep learning methods automatically learn features. Traditional machine learning approaches cannot provide high-level features. They

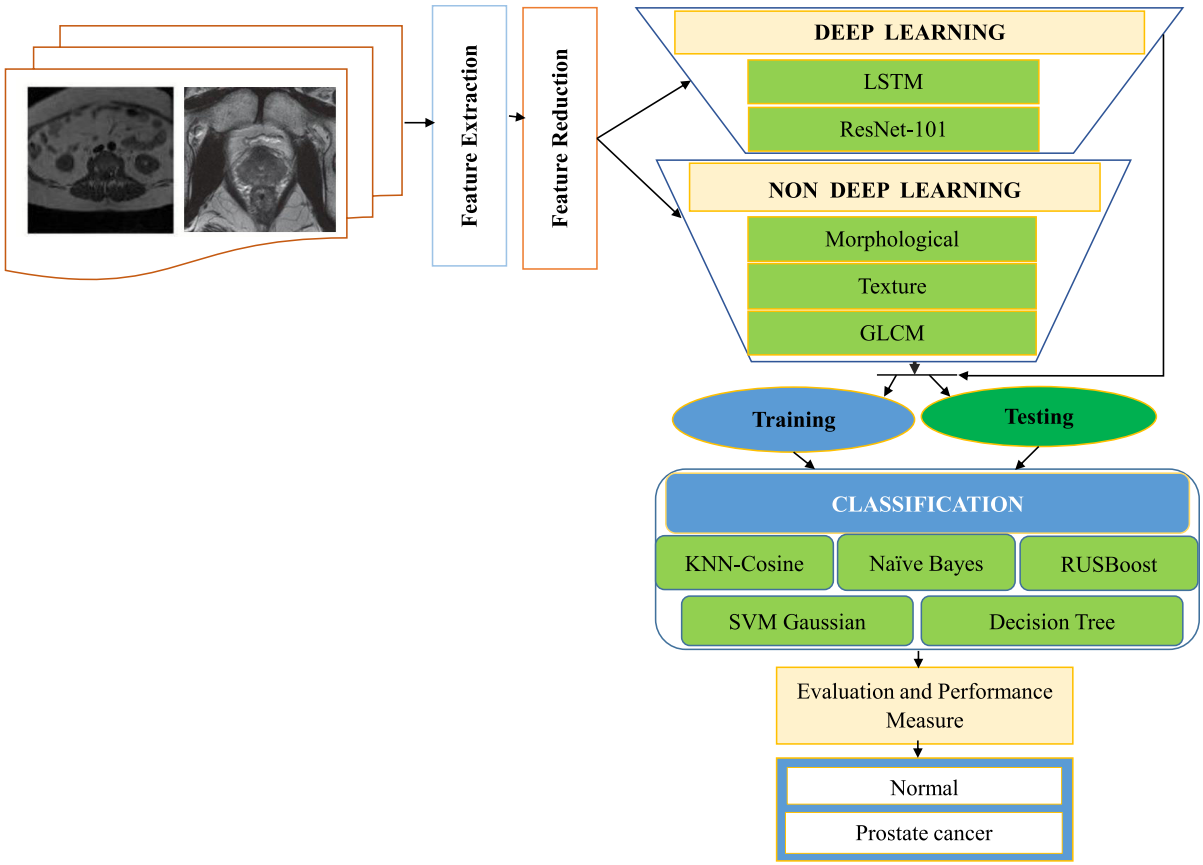

FIGURE 1. Proposed research framework.

need domain expertise where a deep method solves these problems

This study is based on feature extraction, and feature reduction strategy using deep learning and non-deep or machine learning approaches for feature extraction such as morphological features, GLCM, and LSTM, where we have also performed a feature reduction procedure obtained optimal features using autoencoder. We used recent classification techniques that include SVM Gaussian, KNN-Cosine, DT, RUSBoost tree, and Naïve Bayes to classify and detect prostate cancer from Brachytherapy. The aim of this study was to employ and optimize the deep learning methods and comparing the results with non-deep learning methods with the traditional feature extraction approach. We first extracted the texture, morphological and GLCM features from prostate cancer images and then applied machine learning techniques. Among these features, the GLCM features yielded the better detection performance. Then, we optimized and employed the deep learning methods including LSTM and ResNet-101. The LSTM is designed and motivated to control the issues related to the disappearing gradient in the recurrent neural network (*RNN*) architecture. The learning capability of the LSTM structure is quite high, which is supported by the dropout layers against memorization. LSTM framework adds the input barrier, forgetting gate and output gateway towards the neurons in regressive neural network architecture. We then optimized and employed the CNN ResNet-101 with transfer learning approach, which is pre-trained on ImageNet consisted of inception layer, convolution layer and fully connected layer. LSTM yielded higher detection performance than traditional machine learning approaches, while Deep learning ResNet-101 gives the highest detection than LSTM based on its non-linear optimization function, i.e., Rectified Linear Unit (*ReLU*) and back-propagation with gradient descent approach.

Figure 1 presents the overall framework of the proposed study.

II. MATERIAL AND METHODS

A. DATASET

This study uses publicly available data-sets.[1] This dataset contains 230 for MRI scans of patients with different categories and descriptions. This study includes two groups of MRI, the first is the case of the prostate, and the second is Brachytherapy. Also, the last series consists of MRI images of patients used for classification. This dataset was recorded by the Health Insurance Portability and Liability Act 1996 (HIPPA) regulations. An internal review committee also

[1] https://prostatemrimagedatabase.com/

approved it. As part of grant $U41RR019703$, the dataset was funded by the National Institute of Health. The last series contains $T1$ weighted, so all images were diffusion-weighted and dynamic enhanced sequence. This dataset was also reviewed by radiologists unaware of other images, but they examined the prostatectomy. Three pulse sequences were encountered and evaluated by radiologists. The first sequence contains a peripheral zone (PZ), which shows low intensity on $T1$ images and shows the early enhancement. After this lesion, the transitional zone (TZ) is recorded, which also show low intensity for $T1$ images [26], [27]. The current study contains 20 patients with 647 MRI of PCa and 200 MRI of Brachytherapy and 447 MRI from the prostate. For validation of training/test data, 30% of images were selected for validation and testing, and 70% for training.

The Jack-knife 10-fold cross-validation (CV) technique was applied for training/testing data formulation and parameter optimization. The 10-fold CV is the most used method to validate the accuracy of the classifier. Using 10-fold CV, the portioning of the data is made into ten folds; in training, the nine folds participate, and classes of remaining folds are predicted based on the training performed on nine folds. For the trained models, the test samples in the test fold are purely unseen. The entire process is repeated ten times, and each class sample is predicted accordingly. Finally, the unseen samples predicted labels are used to determine the classification accuracy. The process was revised for each parameter; also, classification was reported for samples. K-fold Cross-validation is an effective preventative measure against overfitting. It is a good idea to split the dataset into multiple train-test bins for model tuning. In standard k-fold cross-validation, the data is divided into k-folds using k-fold CV. Iteratively, k-1 folds are involved in the model preparation/training, and the remaining folds are used as the test set. K-fold helps us fine-tune the hyperparameters with the given original training dataset, which helps us determine how an ML model's outcomes could be generalized. The k-fold cross-validation procedure is reflected in Figure 3.

We converted 3D MRI to 2D greyscale images as we prefer greyscale images instead of RGB because a grey image can give us faster processing than images with multiple channels. Yet 3D images are rich in information than 2D; however, it is time-consuming and slow, where 2D images fast, generate more accurate results [28]. 2D images have excellent sensitivity and short image time, best for slow flow. Figure 2 shows a transverse $T2$-weighted fast spin-echo image, obtained 15 months after high-dose-rate brachytherapy for prostate cancer.

B. FEATURE EXTRACTION

Feature extraction is used to capture visual and hidden attributes of images for further processing. These are dimensionality reduction and extracting information from rich data containing medical images, e.g., radiography, MRI, and Computerized Tomography (CT). These modalities contain

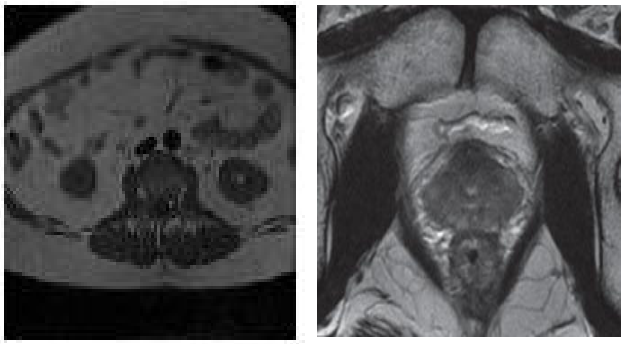

FIGURE 2. Prostate and brachytherapy sample images.

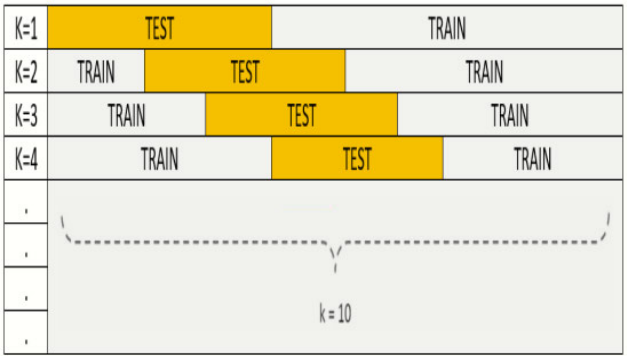

FIGURE 3. The k-fold cross-validation procedure.

different information types, but a feature extraction strategy is used to extract the relevant and desired information.

This study aims to employ and optimize the deep and non-deep learning techniques to improve prostate cancer detection performance. We first extracted texture, morphological and GLCM features and applied non-deep learning techniques; we used some pre-processing steps on texture features to handle the black regions. We then employed deep learning techniques such as LSTM and $ResNet-101$. For LSTM, we optimized the parameters to improve the detection performance. For $ResNet-101$ deep learning method, we first extracted features from $ResNet-101$ from the softmax layer and then applied the feature reduction using autoencoder to use only the most relevant features to enhance the prediction performance. After optimizing and applying the feature reduction methods on ResNet-101, we used robust machine learning techniques to improve the detection performance.

1) TEXTURE BASED FEATURES

A statistically-based feature extraction strategy analyzes how gray image values are spatially distributed. This method is divided into a one-pixel (first-order) and a pixel pair (higher-order). In the one-pixel method, the mean and the variance are used as estimation properties. Mean, median, and mod are also basic statistical methods used in image processing. Higher-order statics estimate properties present at specific

TABLE 1. Mathematical detail of texture features.

Texture Features	Formula
Contrast	$\sum_{i,j=0}^{N-1} p_{i,j}$
Correralation	$\sum_{i,j}^{N-1} p_{i,j}[\frac{(i-\gamma_i)(j-\gamma_j))}{(p_i^2 p_j^2)}]$
Dissimialrity	$\sum_{i,j}^{N-1} p_{i,j}[i,j]$
Energy EN	$\sum_{i,j}^{N-1} p_{ij}^2$
Entropy EP	$\sum_{i,j}^{N-1} p_{i,j}[-\ln p_{i,j}]$
Homogeneity	$\sum_{i,j}^{N-1} \frac{p_{i,j}}{1+(i-j)^2}$
Mean	$\mu_i = \sum_{i,j=0}^{N-1} a(p_{i,j}), \mu_j = \sum_{i,j=0}^{N-1} j(p_{i,j})$
Variance	$\sigma_i^2 \sum_{i,j=0}^{N-1} p_{i,j}(i-\mu_i)^2$, $\sigma_j^2 \sum_{i,j=0}^{N-1} p_{i,j}(j-\mu_j)^2$
Standard deviation	$\sigma_i = \sqrt{\sigma_i^2}$, $\sigma_j = \sqrt{\sigma_j^2}$
Skewness	$\frac{1}{M^2} \sum_{i=0}^{N} \sum_{j=0}^{N} \left(\frac{p(i,j)-\mu}{\sigma}\right)^3$
Kurtosis	$\frac{1}{M^2} \sum_{i=0}^{N} \sum_{j=0}^{N} \left(\frac{p(i,j)-\mu}{\sigma}\right)^4$
RMS	$\sqrt{\frac{p_{i,j}^2}{N^2}}$

locations. In GLCM, several distinct gray levels are equal to the number of rows and columns. It is a matrix that relates the frequency from one gray level to another to specific linear investigating regions of the image. The authors in [29] present GLCM and its distinct features that are the most significant second-order statics. GLCM connects two neighboring pixels in an offset; the gray value is transformed using kernel masks in the co-occurrence matrix. GLCM is a famous second-order feature that describes the relationship between pixels of gray levels. Pre-processing consist of the following two steps

1) calculation of co-occurrences matrix
2) calculation of texture feature using co-occurrences matrix.

GLCM is a square matrix of order M × N, which contains pixels of intensity "x", which is the frequency of event with intensity "y". "x" and "y" are gray-scale calculates how pixels with intensity "x" occurs with pixel "j" at certain distance "d" and and angle θ. GLCM features are used to characterize the pictorial data; its classification can be done based on crop resolution, where the task is to extract meaningful data. While transforming the image to the matrix, different directions are used, such as 0° horizontal, 45° along positive diagonal, 90° vertical, and 135° along negative diagonal. Table 1 contains texture features used in this study.

2) MORPHOLOGICAL FEATURES

Functional details, physiological information, and description (shape and size) of a cell can be obtained using morphological features. Shape and size can be obtained using radius, perimeter, roundness, and compactness. Morphological features have recently been used to localize prostate cancer, cancer cell detection, and colon cancer detection. Here, we have extracted fifteen morphological features listed in Table 2 to extract prostate cancer from MRI images.

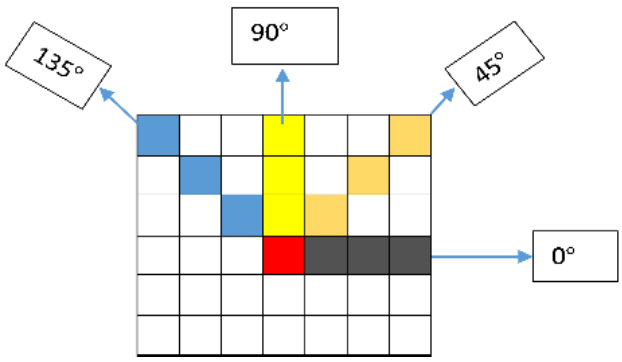

FIGURE 4. GLCM matrix structure at four angle directions.

Figure 4 present the GLCM matrix obtained by transforming an image. This transformation of the image into a

TABLE 2. Mathematical detail of morphological features.

Morphological Features	Equation/Description	Use to measure
Area (A)	Total pixels	Total pixels
Perimeter (P)	Boundary pixels	Total Pixels
Maximum radius (Max r)	Highest distance between centre and outermost boundary	Difference between center and boundary
Minimum radius (Min r)	Lowest distance between innermost and outermost boundary	Minimum distance
Euler Number (Eu No)	Effected – total unaffected region	Difference b/w affected and unaffected
Eccentricity (Ect)	$\sqrt{1-\left(\frac{min.r}{max.r}\right)^2}$	Lengthening
Equiv-Diameter (Ecd)	$\sqrt{4\times area/\pi}$	Diameter of circle
Elongatedness (Egn)	$\left(\frac{area}{2\times max.r}\right)$	Ratio of length to thickness
Entropy (En)	$\sum(p\times \log_2(p)^2)$	Randomness of image
Circularity 1(C1)	$\sqrt{\frac{area}{\pi\times max.r^2}}$	Resemblance with circle
Circularity 2(C2)	$\sqrt{\frac{minr}{maxr}}$	Resemblance with ellipse
Compactness	$\frac{2\times\sqrt{area\times\pi}}{perimeter}$	Deviation from perfect image
Dispersion(DISP)	$\frac{max.r}{area}$	Irregularity
Standard Deviation (STD)	$\frac{1}{k}\sum_{j=1}^{k}(a_j-\overline{a})^2)$	Average contrast
Shape Index (SIN)	$\frac{1}{k}\sum_{j=1}^{k}(a_j-\overline{a})^2)$	Curved images

normalized matrix can be computed as.

$$\begin{aligned}(a,b,d,0^\circ) &= \#((c,d),(e,f)\epsilon(L_j,L_i)\times(L_j,L_i)\\ &\quad |c-e=0|,|d-f=d|)\\ &\quad imj(c,imj)=a,imj(e,f)=b \qquad (1)\\ (a,b,d,45^\circ) &= \#((c,d),(e,f)\epsilon(L_j,L_i)\times(L_j,L_i)\\ &\quad |c-e=d|,|d-f=d|)\\ &\quad imj(c,imj)=a,imj(e,f)=b \qquad (2)\\ (a,b,d,90^\circ) &= \#((c,d),(e,f)\epsilon(L_j,L_i)\times(L_j,L_i)\\ &\quad |c-e=d|,|d-f=0|)\\ &\quad imj(c,imj)=a,imj(e,f)=b \qquad (3)\\ (a,b,d,135^\circ) &= \#((c,d),(e,f)\epsilon(L_j,L_i)\times(L_j,L_i)\\ &\quad |c-e=-d|,|d-f=-d|)\\ &\quad imj(c,imj)=a,imj(e,f)=b \qquad (4)\end{aligned}$$

3) FEATURE REDUCTION

When there are many features in the assessment model, feature reduction is a core concept. it reduces size, reduces processing time, and results in the generalization model [30]. Two major approaches to features reduction are feature extraction and feature selection. The feature selection method selects a few features from the original carpus. But these methods, which use filtering, wrapping, and embedding approaches, may discard or neglect some vital information [31]. On the other hand, shallow feature extraction methods fail to discover relevant detail when the size of data increases.

Feature reduction is a pre-processing step; many algorithms and methods are used for feature reduction; however, some algorithms ignore original information and cannot recover original data. In this study, we have employed an autoencoder (AE) through which we reduce features and reconstruct original data Deep learning has achieved the state of the art performance in image analysis. Autoencoder is a type of advanced neural network. It has many applications one of them is feature reduction. [32] use autoencoder for dimensionality reduction. AE is composed of two parts. The first part is an encoder, while the second part is a decoder. If we have data D with S samples having features f, then output O of an encoder is tunned in such a way that its original dataset D can be obtained from O by minimizing the difference between D and D. Simply it is a function that maps D to O.

$$O=f(D)=a_f(M_D+b_D) \qquad (5)$$

Here a_f is the activation function for encoding. If this linear function is identity, then a_f will perform linear projection, this process is done by encoder function now decoder function g maps hidden outputs to D.

$$D=f(O)=a_g(_D+b_O) \qquad (6)$$

a_g is activation function for decoding. Parameter findings are done by training the auto-encoder $\theta=(M,b_D,b_O)$ which minimize loss objection function.

$$\phi=min_\theta k(D,D)=min_\theta k(D,g(a(D))) \qquad (7)$$

Linear loss construction $L_l\theta$.

$$L_l\theta=\sum_{j=1}^{n}\parallel d_j-(d_j)\parallel^2=\sum_{j=1}^{n}\parallel d_j-g(a((d_j))\parallel^2 \qquad (8)$$

Similarly, for nonlinear loss reconstruction $L_n\theta$

$$L_n\theta=\sum_{j=1}^{n}d_j\log(o_j)+(1-d_j)\log(1-o_j) \qquad (9)$$

Here $d_j\in D(d_j)\in D$ *and* $o_j\in O$

C. MACHINE LEARNING METHODS

This study is based on different non-deep learning or machine learning algorithms to diagnose prostate and brachytherapy cancer from MRI medical images.

1) SUPPORT VECTOR MACHINE

SVM is constructed over discriminant features to predict labels. The discriminant approach is used in multidimensional feature space and posterior probabilities. The authors in [33] proposed SVM. SVM was compared to different algorithms, and the results generated by SVM were more significant than the other artificial intelligence (AI) techniques. SVM classification has also been used for the training and classification of prostate images.

A classifier that generates maximum separation is termed as firm margin or SVM with linear Kernel. When the data is nonlinear instead of a firm margin, SVM correctly classifies the data by widening its margin or behaving according to the data, this procedure is called a soft margin. Using soft margin, SVM transforms data into high dimensional space. Hyperplane can be defined as:

$$g(x) = W \times X + b = 0 \tag{10}$$

Here W is normal to plan. The equation for linearly separable data can be written as for k $= 1, 2 \ldots\ldots N$.

$$(X_k, Y_k) \; where \; X_k \epsilon R^l d \; and \; Y_k \epsilon \{-1, 1\} \; two \; binary \; classes \tag{11}$$

$$W \times X_k + b \geq 1 \; for Y_k = +1 \tag{12}$$

$$W \times C + b \leq 1 \; for Y_k = -1 \tag{13}$$

After combination, both can be written as

$$Y_k(W + b \times X_k) \geq 1 \tag{14}$$

Sign function can be written as

$$g(x) = (W \times X + b) \tag{15}$$

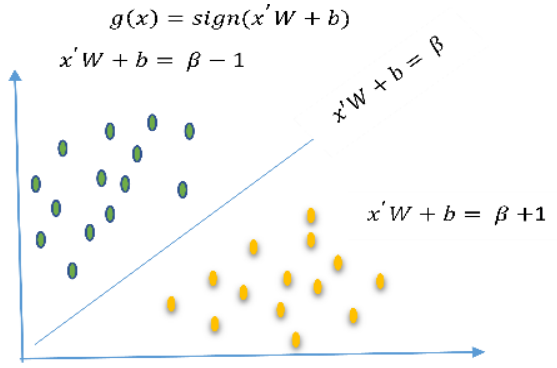

FIGURE 5. Representation of SVM classifier for linear data.

Figure 5 shows linear classification, here β is hyperplane normal vector, and b is offset. In Figure 6, red lines represent decision boundaries by $W \times X + b = 0$, left and right sides represent separate classes. Kernel-based algorithms also

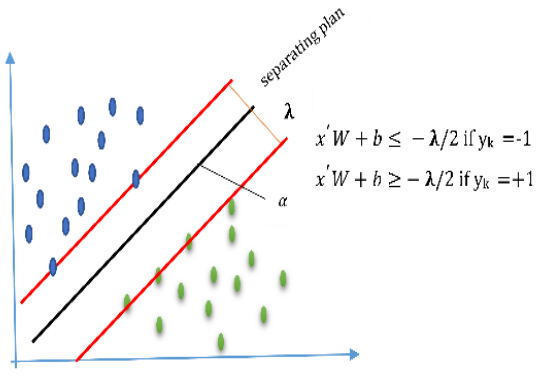

FIGURE 6. Representation of SVM classifier for nonlinear data with decision boundaries.

perform better clustering. Different parameters for SVM, e.g., tolerance (type = *float*, where default value is $1e-3$), verbose (type *bool* default value *False*), max-iteration (type *int* default value is -1), break-ties (type *bool* default value is *False*) and random-state (default value *none*) were used. This study has also used the multi-scale kernel, city block distance matrix and box constraint level. SVM polynomial kernel is represented by

$$P(x_j, y_j) = (x_j, y_{j+1})^n \tag{16}$$

Here p is polynomial Kernel, where n is a tuneable perimeter to manage discriminant function. Similarly, Gaussian RBF kernel is denoted as:

$$P(x_j, y_j) = exp\left(-1/2\frac{(x_j - y_j^2)}{\delta^2}\right) \tag{17}$$

Here δ is smoothing parameter control width of curve. SVM fine Gaussian RBF kernel is defined as

$$P(x_j, y_j) = exp\left(-1/2\frac{x_j - y_j' \| x_j - y_j \|}{\delta^2}\right) \tag{18}$$

2) KNN-COSINE

KNN-Cosine proposed in [34] this is best method used for different problem of classification are used.

$$Cos(a_k, a_l) = 1 - \frac{d_k}{r_{k,l}} = 1 - \frac{d_k}{\sqrt{d_k} + d_l} \tag{19}$$

$$Cos(a_l, a_K) = 1 - \frac{d_k}{r_{k,l}} = 1 - \frac{d_l}{\sqrt{d_l} + d_k} \tag{20}$$

here d is distance of training to test data.$Cos(a_k, a_l)$ and $Cos(a_l, a_K)$ represent cosine weight of nearest neighbour.

$$Sum(D, C) = \sum_{k=1}^{n} cos(a_k, a_l) | C_k = C_r \tag{21}$$

here D is dataset and C is class. For KNN, different distance matrics, Neighbours, inverse square distance weights and neighbors weights were computed.

3) ENSEMBLE RUSBOOST TREE

A decision tree can be constructed using three methods: boosting, bootstrap aggregation, and random forests [35]. Boosting is a popular model-guided selection that is a generic model for the weak learner as a DT. To achieve better accuracy, vulnerable learners combine all classifiers into a single robust composite classifier. Freund, Schapire, and Hill introduced adaptive Boosting or AdaBoost in [36]. The ensemble idea has its origins in supervised machine learning at the late 17^{th} century [37]. The Ensemble method can be used to improve the quality of unsupervised tasks. The ensemble method includes the following blocks

1) Training set: Ensemble training uses a labeled dataset; the input contains a set of attributes.
2) Base inducer consists of an induction algorithm that takes input from a training set and develops a classifier, showing the relationships between input and output attributes.
3) Diversity generation: Used to generate a diverse classifier
4) Combiner: It combines different classifiers and differentiates dependent and independent framework.

In this study, Learning rates and ensemble methods were evaluated for ensemble classifiers.

4) DECISION TREE

A rooted tree is used to classify the basics of recursive partition. It can be combined with other classifiers to improve quantitative examination and predictions [38]. We have tuned Decision tree parameters based on split and maximum split criteria. DT can be constructed using the following equations.

$$\bar{y} = \{y_1, y_2, y_3 \ldots . y_m\}^T \tag{22}$$

$$y_k = \{y_1, y_2, y_3 \ldots . y_{kl} \ldots y_{kn}\} \tag{23}$$

$$p = \{p, p_2, p_3 \ldots . p_k \ldots p_m\} \tag{24}$$

5) KERNEL NAIVE BAYES

The Naive Bayes classifier is robust for many linear classifications. Recently Al-khurayji and Sameh in [39] use its latest Kernel function to classify Arabic text, which shows very efficient results. It has been used to classify interface and non-interface base surfaces. Naïve Bayes is one of the top 10 machine learning algorithms [40]. It is considered because of its attributes, independence, nature, and computational overhead due to attributes selection. Let Y be the measurement vector belonging to the class j.

$$y = (y_1, y_2, y_3 \ldots \ldots y_N) \tag{25}$$

so probability is:

$$P(j_1, j, j_3 \ldots \ldots, y_) = p(S) \prod_{k=1}^{n} p(j_k/y) \tag{26}$$

$$p\left(\frac{j_k}{y_k}\right) = \frac{1}{M_k} \sum_{v=1}^{M} kernal(y_k, y_{vk}),$$

$$kernal(c, d) = \frac{1}{\sqrt{2\pi}} e^{\frac{(c-d)^2}{2n^2}} \tag{27}$$

D. DEEP LEARNING METHODS

Artificial intelligence (AI) works the same way as the human brain. Deep learning is part of artificial intelligence. Like a human brain, Artificial Neural Networks (ANNs) take information and process that information using a group of neurons to form layers. These neurons transfer information to other neurons where certain information is fed back to the previous layer. Finally, the processed information is sent to the output layer in the form of classification or regression. Deep learning methods automatically extract features from data to improve the prediction of complex problems [41].

1) LONG SHORT-TERM MEMORY (LSTM)

Hochreiter and Schmidhuber proposed the LSTM algorithm in 1997 [42]. Alex Graves improves LSTM, which solves the gradient disappearance problem. LSTM improves gait recognition in degenerate neural disease compared to older methods. In the machine learning approach, gradient learning remains a problem. LSTM solves this problem because it is based on an appropriate gradient-based learning algorithm that solves the error backflow problem. LSTM generates optimal results even with noise and incompressible input sequence without losing short time lag capabilities.

The authors in [42] proposed a deep LSTM-RNN; this method uses an end-to-end learning procedure that combines different LSTM-RNN to solve a vanishing gradient problem. The hidden layer in LSTM is linear, but self-loop memory blocks allow the gradient to flow through large sequences. LSTM consist of recurring blocks called memory blocks. The memory chip consists of blocks of different versions. Each block contains recurring memory cells and three multiplicative units input, output, and forget gates [43]. These cells allow memory blocks to store and access information for a long time to solve the vanishing gradient problem [42].

LSTM rollback is an optimized combination of back propagation, throughput, and real-time recurrent learning. The output unit gate and forget gate use the error-free version of real-time recurrent learning (RTRL).

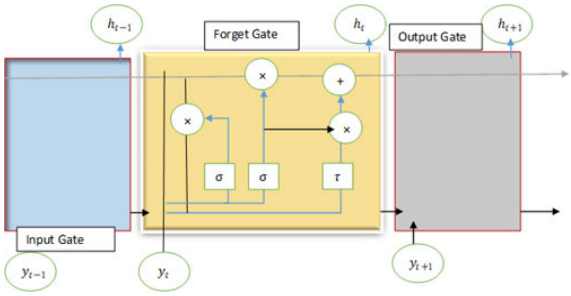

FIGURE 7. LSTM model based on recurrent neural network.

Figure 7 shows the LSTM model based on the recurrent neural network, which is composed of the input gate that controls information flow, the forget gate that determines the amount of information erased from memory, the output gate that is responsible for controlling the output with

four operations (three sigmoids and one tanh) Before intake, LSTM standardizes the dataset using the following equations.

$$\mu = \frac{1}{k}\sum_{i=0}^{k-1} x_0[i] \tag{28}$$

$$v = \frac{1}{k-1}\sum_{i=0}^{k-1}(x_0[i]-\mu)(x_0[i]-\mu)^M \tag{29}$$

So, data is standardized for training using these equations. After input LSTM forward to next cell.

$$k_s[i] = W_{as}x[i] + W_{ss}s[i-1] + W_{ys}y[i-1] + L_s \tag{30}$$

$$k_o[i] = W_{ao}x[i] + W_{so}s[i-1] + W_{yo}y[i-1] + L_o \tag{31}$$

We create an LSTM model with one LSTM layer of 64 neurons and " RELU" as an activation function. After that we added four dense layers, where first three layers contains 32, 16 and 8 neurons respectively with (*RELU*) activation function. The final layer which also acts as the output layer, contains 1 neuron. Finally, we compiled our model using optimizer = (*ADAM*) and train it for 100 epochs with a batch size of 24.

2) RESIDUAL NETWORK (ResNet-101)

In recent years supremacy of deep Networks has been reported in several works [44]. Residual Network (ResNet-101) was designed by authors in [45], which is the inspiration of [46] that is the deepest architecture of ImageNet. ResNet-101 uses the same number of layers and filters for the same output features. ResNet-101 uses residual connection after application of the chain rule of integration. Residual networks are composed of many selected units, where each unit can be represented as

$$y_i = h(a_i) + F(a_i, W_i) \tag{32}$$

and

$$a_{i+1} = F(y_i) \tag{33}$$

here a_i is input $F(a_i, w_i)$ is the residual function $a_{(i+1)}$ depends on y_i so $a_{(i+1)}$ is output. The Residual map can be formulated as $F = W_i\delta(W_i a)$ here δ denote Relu [47]. if dimensions of a_i and y_i are not equal or if we change input and output channels the linear projection W_m is performed to match dimensions.

$$y_i = F(a_i, W_i) + W_m a \tag{34}$$

square matrix of W_m can also be used in Equation 30 but W_m is only used in matching dimensions. Equation 30 is similar to the linear function $y_i = W_i\, a + a$ but residual function $F(a_i, W_i)$ Represents multiple convolution layers. ResNet-101 can be written in iterative form.

$$x_I = x_i + \sum_{n=i}^{i} -1\, F(a_i, W_i) \tag{35}$$

E. PERFORMANCE MEASURES

1) CONFUSION MATRIX

True positive rate (TPR)

True Positive Rate is also called sensitivity, this shows positively tested subjects under examination. Mathematically, it can be represented as:

TPR = sum of true positive/sum of positive condition

$$TPR = \frac{TP}{TP + FN} \tag{36}$$

True negative rate (TNR)

TNR also called specificity shows negative tested identified correctly. Mathematically, it can be represented as:

TNR = sum of true negative/sum of negative condition)

$$TNR = \frac{TN}{FP + TN} \tag{37}$$

Positive predictive value (PPV)

PPV = sum of true positive/sum of predicted positive

$$PPV = \frac{TP}{TP + FP} \tag{38}$$

Negative predictive value (NPV)

NPV = sum of true Negative/predicted Negative Condation

$$NPV = \frac{TN}{TN + FN} \tag{39}$$

Accuracy (AC)

$$AC = \frac{TP + TN}{TP + FP + FN + TN} \tag{40}$$

2) RECEIVER OPERATING CURVE (ROC)

ROC is a standard method used in the evaluation of classifiers [48]. A portion of a square unit is considered to be the area under the curve. AUC is part of a square area whose value is between 0 and 1. Any realistic classifier's value should be greater than 0.5, where 0.9 to 1.0 means a value of 90% to 100%. ROC is a great test and plotted according to specificity and sensitivity. We take a true positive rate of prostate cancer along the y-axis, where the false positive rate along the x-axis. This study achieved a maximum AUC of 1 using a deep classifier ResNet-101 and 0.9999 using for LSTM features, followed by 0.9984 by KNN-Cosine for GLCM and 0.9974 by SVM Gaussian for GLCM features, Gaussian followed by KNN-Cosine. The maximum separation (AUC) for SVM Gaussian was also 95%.

3) MATHEWS CORRELATION COEFFICIENT (MCC)

Matthews developed the MCC in 1975 for the analysis of chemical structure [49]. It was proposed again by the authors in [50] for machine learning. When the dataset is imbalanced, or one class is more extensive than others, accuracy measures cannot be considered reliable. The practical solution to this problem is MCC. In this study, we employed MCC accuracy along with ROC. The results show that the LSTM has an MCC value of 98.95%,where deep method ResNet-101 has

an MCC value of 1 using Kernel Naive Bayes, SVM Gaussian and RUSBoost Tree classifiers, which outperforms other classifiers.

4) 10 FOLD CROSS VALIDATION

In prediction, estimation is necessary to evaluate the performance of the prediction model [51]. Cross-validation (CV) is a common method of model selection and error prediction that splits data multiple times for risk estimation. CV selects the lowest risk error after dividing the data into training and testing. In this study, we used k-fold cross validation where k = 1, 2,3,4,5, and 10 for each k iteration, the remaining K-1 folds are used for learning. We found that when $k = 10$, it generates a better classification. The results computed in this study depicted 10-fold cross-validation.

III. RESULTS AND DISCUSSION

Deep and non-deep classifiers like SVM radial function, Naive Bayes, RUSBoost tree, and decision tree were used to evaluate the results for different features (morphological, texture, GLCM, and LSTM).

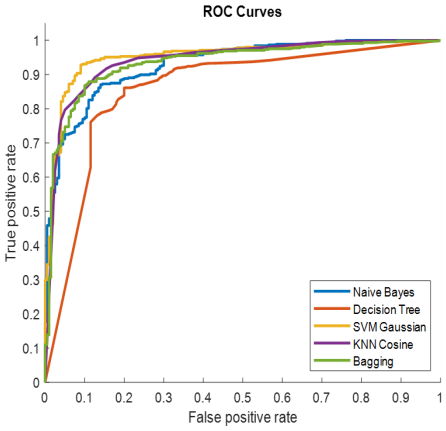

FIGURE 8. ROC using texture features.

Using the machine learning method on image scanning for automatic PCa grading. They obtain 82% ROC accuracy using logistic regression combining with morphological and quantitative information of glands and stroma. The receiver operating curve graph is a methodology used for performance measurement. A single value, AUC, evaluates performance using ROC. We achieved 99.84% ROC using KNN-Cosine, which is higher among all non-deep classifiers using GLCM-based features. Figure 8 compares different classifiers. We can conclude that using texture features, the Gaussian SVM Classifier produces optimal results, where all other non-deep learning methods also had better results.

Figure 9 compares different classifiers using pre-processing texture features. Results depicts that with pre-processing texture features, the Gaussian SVM Classifier and KNN-Cosine produce optimal results, where all other non-deep learning methods also had better results except Kernel Naive Bayes. Results also reveal that Using texture

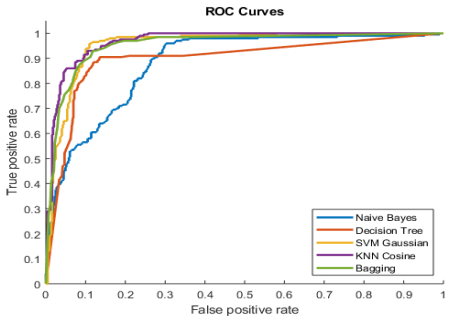

FIGURE 9. ROC using texture features with pre-processing.

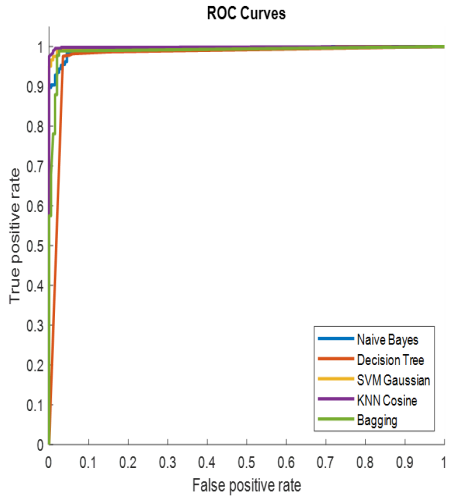

FIGURE 10. ROC using GLCM features.

features only SVM Gaussian produces better results; on the other hand, all classifiers produce better results using pre-processed texture features.

Similarly, Figure 10 also shows that all the classifiers correctly separate the positive and negative cases using GLCM matrix features. The lowest precision was demonstrated by the kernel-based naive Bayes, which is 95.21%. It is implied that all the classifiers achieved a higher classification than the one with the lowest performance. KNN-Cosine produces a better classification than SVM Gaussian. Non-deep learning classifiers were used on the morphological features extracted from MRI images to diagnose prostate cancer. Different measures of assessment of the confusion matrix were also applied to these classifiers to evaluate the results.

The results in Table 3 shows that using different machine learning classifiers on texture features, the highest sensitivity was obtained by Kernel Naïve Bayes (89.00%), followed by SVM Gaussian (85.50%). Similarly, SVM Gaussian achieved maximum specificity (94.18%), maximum accuracy (91.50%), and maximum separation as AUC (0.9533). Moreover, the highest NPV was obtained by Kernel Naïve Bayes (94.39%) followed by the RUSBoost tree and then SVM Gaussian, where the highest PPV was obtained by SVM. We computed texture features using pre-processing

TABLE 3. Summary of texture based features.

Classifier	Sens	Spec	PPV	NPV	Accuracy	FPR	AUC	MCC
Kernel Naïve Bayes	0.895	0.7897	0.6557	0.9439	0.8223	0.2103	0.9322	0.6407
Decision Tree	0.6900	0.8971	0.7500	0.8661	0.8331	0.1029	0.8517	0.6014
SVM Gaussian	0.8550	0.9418	0.8680	0.9356	0.9150	0.0581	0.9533	0.8001
KNN-Cosine	0.850	0.9083	0.8057	0.9312	0.8903	0.0917	0.9469	0.7475
RUSBoost Tree	0.870	0.8926	0.7838	0.9388	0.8856	0.1074	0.9432	0.7423

TABLE 4. Summary of texture based features with pre-processing.

Classifier	Sens	Spec	PPV	NPV	Accuracy	FPR	AUC	MCC
Kernel Naïve Bayes	0.820	0.7651	0.6097	0.9048	0.7821	0.2349	0.8813	0.5486
Decision Tree	0.805	0.9128	0.805	0.9128	0.8794	0.08725	0.8969	0.7178
SVM Gaussian	0.856	0.9128	0.816	0.9379	0.898	0.08725	0.9547	0.7658
KNN-Cosine	0.86	0.9463	0.8776	0.9379	0.9196	0.05369	0.9674	0.8109
RUSBoost Tree	0.845	0.9262	0.8366	0.9303	0.9011	0.07383	0.9539	0.7691

TABLE 5. Summary of GLCM based features.

Classifier	Sens	Spec	PPV	NPV	Accuracy	FPR	AUC	MCC
Kernel Naïve Bayes	0.9600	0.9485	0.8930	0.9815	0.9521	0.05145	0.9945	0.8913
Decision Tree	0.9650	0.9754	0.9461	0.9842	0.9722	0.02461	0.9753	0.9353
SVM Gaussian	0.9700	0.9911	0.9798	0.9866	0.9845	0.00894	0.9975	0.9637
KNN-Cosine	0.9800	0.9955	0.9899	0.9911	0.9907	0.004474	**0.9984**	0.9782
RUSBoost Tree	0.9950	0.9754	0.9476	0.9977	0.9815	0.02461	0.9835	0.9577

steps in Table 4, results depicts that highest sensitivity (86.00%) was achieved by *KNN − Cosine* followed by *SVMGaussian* (85.60%), where maximum specificity (94.63%) was obtained using *KNN − Cosine*, by the procedure of pre-processing we achieve accuracy of (91.96%) and AUC value as (0.9674).

Table 5 shows that by extracting GLCM features and employing different machine learning classifiers, the highest accuracy was obtained by *KNN − Cosine* (99.07%), followed by SVM Gaussian and RUSBoost tree (98.15%). Maximum separation or larger AUC (0.998) was obtained using the KNN-Cosine classifier, followed by SVM Gaussian Kernel Naïve Bayes (99.45%). Similarly, SVM Gaussian has the highest MCC (0.7540) for texture features. With GLCM features, the highest sensitivity (99.50%) was achieved by RUSBoost Tree, while the highest specificity (99.55%) was achieved using the KNN-Cosine classifier. In the case of positive predictive value, KNN-Cosine obtains PPV (98.99%), and SVM Gaussian obtain PPV (97.98%), followed by the RUSBoost tree having PPV (94.76%) and the tree decision obtained PPV (94.6%) respectively. We have reached a sensitivity (99.50%). The results show that the machine learning classifiers KNN-Cosine, SVM Gaussian, RUSBoost Tree, Decision Tree, and Kernel Naïve Bayes perform better for GLCM features, compared to traditional texture features.

Table 6 describes the results obtained using these artificial intelligence-based approaches (deep learning and non-deep learning). The non-deep learning method was used on morphological features, where deep learning methods were used to classify *LSTM* features extracted from the same images. The results show that KNN-Cosine achieved better accuracy of 989.95%) compared to Gaussian SVM (89.64%). Similarly, the sensitivity and specificity obtained by KNN-Cosine were (83%) and (93%), where SVM Gaussian reached (80%) and (93.9%), respectively. Moreover, the maximum separation obtained by KNN-Cosine is 0.966, where SVM Gaussian has an AUC value equal to (0.945). Similarly, for imbalance data, the MCC is a better evaluation matrix. KNN-Cosine obtained (0.7638) MCC values where SVM Gaussian obtained (0.7540) MCC values. It can be seen that by using decision tree and ensemble tree classifiers, the accuracy of (87.94%) was achieved by the RUSBoost tree, where DT reaches the accuracy (86.24%) to classify the prostate and Brachytherapy using morphological features of MRI images. RUSBoost has a higher sensitivity, PPV, NPV, and FPR than DT, where DT has a higher specificity value (91.9%).

Using deep learning classifiers on MRI-extracted LSTM features to detect prostate cancer, it is evident that we have achieved (100%) specificity and positive predictive value, where the accuracy (99.48%) was obtained using the same

TABLE 6. Comparison of morphological and LSTM feature.

Classifier	Sens	Spec	PPV	NPV	Accuracy	FPR	AUC	MCC
Kernel Naïve Bayes	0.9650	0.7785	0.661	0.9803	0.8362	0.2215	0.931	0.6905
Decision Tree	0.7350	0.9195	0.8033	0.8858	0.8624	0.0805	0.8934	0.6716
SVM Gaussian	0.8000	0.9396	0.8556	0.913	0.8964	0.0604	0.945	0.7540
KNN-Cosine	0.8300	0.9306	0.8426	0.9244	0.8995	0.0693	0.9665	0.7638
RUSBoost Tree	0.845	0.8949	0.7824	0.9281	0.8794	0.1051	0.9461	0.7225
LSTM	0.9833	1	1	0.9926	0.9948	0	0.9999	0.9879

TABLE 7. ResNet-101 based features.

Classifier	Sens	Spec	PPV	NPV	Accuracy	FPR	AUC	MCC
Kernel Naive Bayes	1	1	1	1	1	0	1	1
Decision Tree	0.995	1	1	0.9978	0.9985	0	0.9975	0.9964
SVM Gaussian	1	1	1	1	1	0	1	1
KNN-Cosine	1	0.9978	0.995	1	0.9985	0.002237	1	0.9964
RUSBoost Tree	1	1	1	1	1	0	1	1

technique. The deep learning classifiers ResNet-101 achieved the highest AUC value of (1) and MCC (1). The results also show that the LSTM classifier and ResNet-101 has a false positive rate of (0%), which is the lowest among all classifiers.

Overall results reveal that GLCM features provide better results for non-deep classifiers. KNN-Cosine (99%) accuracy, MCC (0.9782) And AUC (0.998) among texture features. Similarly, LSTM features extracted from the prostate are more robust, and using deep learning, we have produced optimal results as (99.48%) with (100%) specificity and (100%) positive predicted value (0.9879) MCC values and (0.9999) AUC values,where ResNet-101 produced 100%) specificity and (100%) positive predicted value (1) MCC values and (1) AUC values using Kernel Naive Bayes,SVM Gaussian and RUSBoost Tree which are best among all classifiers.

Similarly, using ResNet-101, we have obtained all optimal results than other machine and deep learning classifiers table Table 7 depicts the result of all classifiers using ResNet-101 approach only DT produce (0.995%) sensitive, (0.9978%) NPV and (0.9985%) AC. AUC value is (0.9975) where other classifiers generate (100%) AC and AUC value as (1), which is better than all machine-based classifiers and LSTM.

The Figure 11 shows the accuracy and loss curve graph using LSTM. The training accuracy represents the classification accuracy on each mini-batch. The light blue line represents the training accuracy. The smooth training accuracy is denoted by the blue color obtained by applying the smoothing algorithm to the training accuracy. The smoothed training accuracy is less noisy than the unsmooth accuracy making it easier to spot trends. After 10th epoch, the highest accuracy of 99.48% was obtained. The next figure shows the loss function against each iteration. The light red color denotes the loss, and

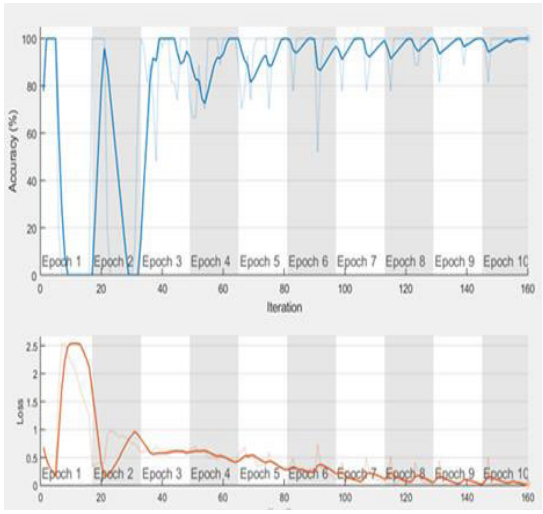

FIGURE 11. Accuracy and Loss function Curves.

the red color indicates the smoothed loss. The smallest loss value was obtained at 10th epoch.

Figure 12 shows the results obtained through morphological features using a machine learning approach. KNN-Cosine performs better than other classifiers. Similarly, Figure 13 indicates that the LSTM deep learning approach outperforms different machine learning classifiers. Figure 14 depicts that mostly AUC is similar, so AUC line colors with equal values are mixed. AUC value is maximum for all classifiers, as all classifiers achieve an AUC value of 1 except DT, which has 0.9975. The red color in the figure indicates DT where all other classifiers have a maximum AUC of 1 so, only the last occurred color, green, can be seen in the figure.

Table 8 presents the comparison of different machine learning and deep learning methods with the proposed study.

TABLE 8. Comparison of proposed study with previous studies.

Reference	Problem	Approach	Performance
[52]	Early diagnosis of PCa	ML: SVM and ANN	AC 79%
[53]	PCa Diagnosis	ML: ANN	AC 94.44%
[54]	Improvement of PCa detection through segmentation	ML: ANN, analysis of MR Spectroscopic imaging	AC 94.9 %
[55]	PCa detection improvement analysis	ML analysis of MR	AC 98.3%
[56]	Gleason grading of PCa	ML quantitative phase imaging	ROC 82%
[57]	Locating and Predicting tumour in PCa tissue	ML on gene expression data	AC 99.0 %
[14]	PCa detection	ML, DL and CNN with transfer learning	AC 99.1 % AUC 0.9999
Proposed study	PCa Detection using deep and non-deep learning techniques	ML: SVM (hand crafted) DL: LSTM,ResNet (non-hand crafted)	AC 100 Spec 100% MCC 100

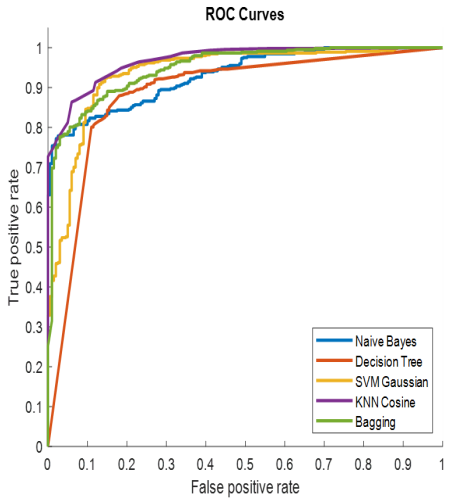

FIGURE 12. ROC using morphological features.

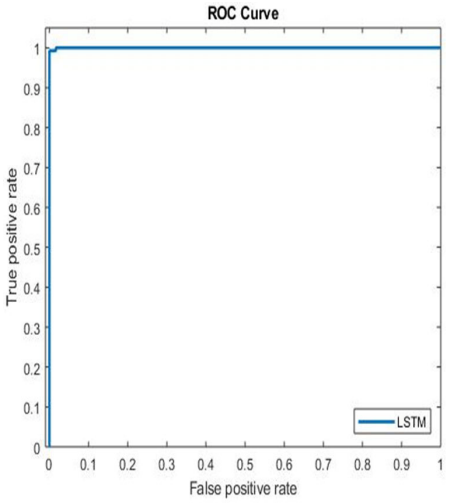

FIGURE 13. ROC using LSTM features.

Linear SVM with ANN was used by [52] to detect prostate cancer and obtained an accuracy of (79%). The author in [53] use ANN to detect PCa and achieve an accuracy of (94.44%). Similarly, authors in [54] use segmentation and ANN procedure to detect prostate cancer; they achieve

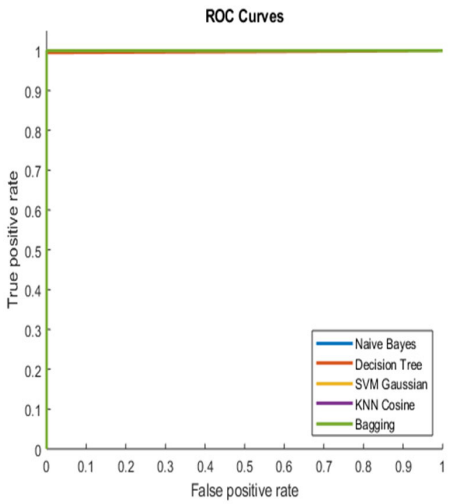

FIGURE 14. ROC using ResNet-101 features.

accuracy of (94.90%). The results show that using a machine learning approach, the prediction of tumors in the prostate using gene expression data was achieved with (99.0%) accuracy. Additionally, deep learning also classifies normal and PCa subjects with (99.1%) accuracy. In this study, we used machine learning and deep learning techniques. The results reveal that our proposed methods produced optimal results.

IV. CONCLUSION

This paper presented an approach to extract features from prostate images using a 10-fold cross-validation test on these hand-crafted and non-hand-crafted features. Features reduction using autoencoder is also done to optimize the features. We did pre-processing on texture features, and The performance was evaluated using confusion matrix, true positive rate (sensitivity), true negative rate (specificity), positive predictive value, negative predictive value, and AUC. We achieved better performance accuracy (99.07%) and AUC (0.9984) with KNN-Cosine for GLCM features using machine learning classifiers. Optimal results were obtained using ResNet-101 and Deep Learning LSTM, i.e., accuracy (99.84%), AUC (0.9999), MCC (0.98.79), and (100%)

specificity. Results depict that the pre-processing process generates better results. The results also conclude that non-handed LSTM fine-tuned tools are more robust than the handed feature extraction strategy.

5) LIMITATIONS OF STUDY AND FUTURE RECOMMENDATIONS

LSTM behaves like feed-forward, and bit parity can be an issue with LSTM. In the future, we will apply LSTM and other deep learning methods with transfer learning approaches, feature selection, and ranking methods to classify prostatic and non-prostatic subjects. We will also obtain the clinical profiles of the patients and compare the results with the larger dataset. In this study, we did pre-processing with texture features; however, we will apply these steps to other feature extraction methods for future findings.

ACKNOWLEDGMENT

The authors want to express our thanks to Artificial Intelligence and Data Analytics (AIDA) Lab (Prince Sultan University, Saudi Arabia) for their kind support for this research. They would also like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) for this publication. We also appreciate the constructive comments of reviewers.

REFERENCES

[1] K. Sklinda, M. Frnczek, B. Mruk, and J. Walecki, "Normal 3T MR anatomy of the prostate gland and surrounding structures," *Adv. Med.*, vol. 2019, pp. 1–9, May 2019.

[2] K. H. Hammerich, G. E. Ayala, and T. M. Wheeler, *Anatomy of the Prostate Gland and Surgical Pathology of Prostate Cancer*. Cambridge, U.K.: Cambridge Univ. Press, 2009, pp. 1–10.

[3] H. Ideo, J. Kondo, T. Nomura, N. Nonomura, M. Inoue, and J. Amano, "Study of glycosylation of prostate-specific antigen secreted by cancer tissue-originated spheroids reveals new candidates for prostate cancer detection," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, Dec. 2020.

[4] R. A. Mahumud, K. Alam, J. Dunn, and J. Gow, "The burden of chronic diseases among Australian cancer patients: Evidence from a longitudinal exploration, 2007-2017," *PLoS ONE*, vol. 15, no. 2, Feb. 2020, Art. no. e0228744.

[5] N. M. Schultz, K. O'Day, R. Sugarman, and K. Ramaswamy, "Budget impact of enzalutamide for nonmetastatic castration-resistant prostate cancer," *J. Managed Care Specialty Pharmacy*, vol. 26, no. 4, pp. 538–549, Apr. 2020.

[6] M. B. Cook, D. C. Beachler, L. E. Parlett, P. T. Cochetti, W. D. Finkle, S. Lanes, and R. N. Hoover, "Testosterone therapy in relation to prostate cancer in a us commercial insurance claims database," *Cancer Epidemiol. Prevention Biomarkers*, vol. 29, no. 1, pp. 236–245, 2020.

[7] M. Wang, G. Chi, Y. Bodovski, S. L. Holder, E. J. Lengerich, E. Wasserman, and A. C. McDonald, "Temporal and spatial trends and determinants of aggressive prostate cancer among black and white men with prostate cancer," *Cancer Causes Control*, vol. 31, no. 1, pp. 63–71, Jan. 2020.

[8] P. C. Albertsen, J. A. Hanley, D. F. Gleason, and M. J. Barry, "Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer," *Jama*, vol. 280, no. 11, pp. 975–980, 1998.

[9] Y. Peng, Y. Jiang, C. Yang, J. B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. L. Giger, S. E. Eggener, and A. Oto, "Quantitative analysis of multiparametric prostate MR images: Differentiation between prostate cancer and normal tissue and correlation with Gleason score—A computer-aided diagnosis development study," *Radiology*, vol. 267, no. 3, pp. 787–796, Jun. 2013.

[10] M. Moradi, S. E. Salcudean, S. D. Chang, E. C. Jones, N. Buchan, R. G. Casey, S. L. Goldenberg, and P. Kozlowski, "Multiparametric MRI maps for detection and grading of dominant prostate tumors," *J. Magn. Reson. Imag.*, vol. 35, no. 6, pp. 1403–1413, Jun. 2012.

[11] S. Sarkar and S. Das, "A review of imaging methods for prostate cancer detection: Supplementary issue: Image and video acquisition and processing for clinical applications," *Biomed. Eng. Comput. Biol.*, vol. 7, Jan. 2016, Art. no. BECB.S34255.

[12] J. Skowronek, "Current status of brachytherapy in cancer treatment–short overview," *J. Contemp. Brachytherapy*, vol. 9, no. 6, p. 581, 2017.

[13] L. J. Korb and M. K. Brawer, "Modern brachytherapy for localized prostate cancers: The Northwest hospital (Seattle) experience," *Rev. Urol.*, vol. 3, no. 1, p. 51, 2001.

[14] A. A. Abbasi, L. Hussain, I. A. Awan, I. Abbasi, A. Majid, M. S. A. Nadeem, and Q.-A. Chaudhary, "Detecting prostate cancer using deep learning convolution neural network with transfer learning approach," *Cognit. Neurodynamics*, vol. 4, pp. 523–533, Apr. 2020.

[15] J.-K. Park, B.-K. Kwon, J.-H. Park, and D.-J. Kang, "Machine learning-based imaging system for surface defect inspection," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 3, no. 3, pp. 303–310, Jul. 2016.

[16] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," 2015, *arXiv:1503.04069*. [Online]. Available: http://arxiv.org/abs/1503.04069

[17] L. Hussain, A. Ahmed, S. Saeed, S. Rathore, I. A. Awan, S. A. Shah, A. Majid, A. Idris, and A. A. Awan, "Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies," *Cancer Biomarkers*, vol. 21, no. 2, pp. 393–413, Feb. 2018.

[18] L. Hussain, S. Saeed, I. A. Awan, A. Idris, M. S. A. Nadeem, and Q.-U.-A. Chaudhry, "Detecting brain tumor using machines learning techniques based on different features extracting strategies," *Current Med. Imag. Formerly Current Med. Imag. Rev.*, vol. 15, no. 6, pp. 595–606, Jul. 2019.

[19] K. Yesu, H. J. Chakravorty, P. Bhuyan, R. Hussain, and K. Bhattacharyya, "Hybrid features based face recognition method using artificial neural network," in *Proc. 2nd Nat. Conf. Comput. Intell. Signal Process. (CISP)*, Mar. 2012, pp. 40–46.

[20] S. Yoo, I. Gujrathi, M. A. Haider, and F. Khalvati, "Prostate cancer detection using deep convolutional neural networks," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Dec. 2019.

[21] S. Liu, H. Zheng, Y. Feng, and W. Li, "Prostate cancer diagnosis using deep learning with 3D multiparametric MRI," *Proc. SPIE*, vol. 10134, Mar. 2017, Art. no. 1013428.

[22] R. U. Khan, X. Zhang, and R. Kumar, "Analysis of ResNet and GoogleNet models for malware detection," *J. Comput. Virol. Hacking Techn.*, vol. 15, no. 1, pp. 29–37, Mar. 2019.

[23] P. Ghosal, L. Nandanwar, S. Kanchan, A. Bhadra, J. Chakraborty, and D. Nandi, "Brain tumor classification using ResNet-101 based squeeze and excitation deep neural network," in *Proc. 2nd Int. Conf. Adv. Comput. Commun. Paradigms (ICACCP)*, Feb. 2019, pp. 1–6.

[24] S. Nitta, M. Tsutsumi, S. Sakka, T. Endo, K. Hashimoto, M. Hasegawa, T. Hayashi, K. Kawai, and H. Nishiyama, "Machine learning methods can more efficiently predict prostate cancer compared with prostate-specific antigen density and prostate-specific antigen velocity," *Prostate Int.*, vol. 7, no. 3, pp. 114–118, Sep. 2019.

[25] R. Alkadi, F. Taher, A. El-baz, and N. Werghi, "A deep learning-based approach for the detection and localization of prostate cancer in t2 magnetic resonance images," *J. Digit. Imag.*, vol. 32, no. 5, pp. 793–807, Oct. 2019.

[26] L. Lemaitre, P. Puech, E. Poncelet, S. Bouyé, X. Leroy, J. Biserte, and A. Villers, "Dynamic contrast-enhanced MRI of anterior prostate cancer: Morphometric assessment and correlation with radical prostatectomy findings," *Eur. Radiol.*, vol. 19, no. 2, pp. 470–480, Feb. 2009.

[27] A. L. Chesnais, E. Niaf, F. Bratan, F. Mège-Lechevallier, S. Roche, M. Rabilloud, M. Colombel, and O. Rouvière, "Differentiation of transitional zone prostate cancer from benign hyperplasia nodules: Evaluation of discriminant criteria at multiparametric MRI," *Clin. Radiol.*, vol. 68, no. 6, pp. e323–e330, Jun. 2013.

[28] R. Usha and K. Perumal, "SVM classification of brain images from MRI scans using morphological transformation and GLCM texture features," *Int. J. Comput. Syst. Eng.*, vol. 5, no. 1, pp. 18–23, 2019.

[29] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[30] X. Yin and Y. Liu, "Deep learning based feature reduction for power system transient stability assessment," in *Proc. TENCON-IEEE Region 10th Conf.*, Oct. 2018, pp. 2308–2312.
[31] S. Ye, X. Wang, Z. Liu, and Q. Qian, "Dual-stage feature selection for transient stability assessment based on support vector machine," *Proc. CSEE*, vol. 30, no. 31, pp. 28–34, 2010.
[32] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, Apr. 2016.
[33] A. Sohail and F. Arif, "Supervised and unsupervised algorithms for bioinformatics and data science," *Prog. Biophys. Mol. Biol.*, vol. 151, pp. 14–22, Mar. 2020.
[34] E. Prasetyo, R. Purbaningtyas, and R. D. Adityo, "Cosine K-nearest neighbor in milkfish eye classification," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 3, pp. 11–25, Jun. 2020.
[35] S. Mounce, K. Ellis, J. Edwards, V. Speight, N. Jakomis, and J. Boxall, "Ensemble decision tree models using RUSBoost for estimating risk of iron failure in drinking water distribution systems," *Water Resour. Manage.*, vol. 31, no. 5, pp. 1575–1589, 2017.
[36] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. Int. Conf. Mach. Learn.*, vol. 96, 1996, pp. 148–156.
[37] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, nos. 1–2, pp. 1–39, Feb. 2010.
[38] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, "Classification of Tweets data based on polarity using improved rbf kernel of SVM," *Int. J. Inf. Technol.*, pp. 1–16, 2020, doi: 10.1007/s41870-019-00409-4.
[39] R. Al-Khurayji and A. Sameh, "An effective Arabic text classification approach based on kernel Naive Bayes classifier," *Int. J. Artif. Intell. Appl.*, vol. 8, no. 6, pp. 1–10, Nov. 2017.
[40] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip, and Z. H. Zhou, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
[41] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[42] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling With Recurrent Neural Networks*. Berlin, Germany: Springer-Verlag, 2012, pp. 37–45.
[43] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.
[44] G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2924–2932.
[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[47] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 1–8.
[48] K. Hajian-Tilaki, "Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation," *Caspian J. Internal Med.*, vol. 4, no. 2, p. 627, 2013.
[49] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochimica et Biophysica Acta (BBA)-Protein Struct.*, vol. 405, no. 2, pp. 442–451, Oct. 1975.
[50] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: An overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, May 2000.
[51] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc., Ser. B (Methodol.)*, vol. 36, no. 2, pp. 111–133, Jan. 1974.
[52] M. Çınar, M. Engin, E. Z. Engin, and Y. Z. Ateşçi, "Early prostate cancer diagnosis by using artificial neural networks and support vector machines," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6357–6361, Apr. 2009.
[53] I. Saritas, I. A. Ozkan, and I. U. Sert, "Prognosis of prostate cancer by artificial neural networks," *Expert Syst. Appl.*, vol. 37, no. 9, pp. 6646–6650, Sep. 2010.
[54] L. Matulewicz, J. F. A. Jansen, L. Bokacheva, H. A. Vargas, O. Akin, S. W. Fine, A. Shukla-Dave, J. A. Eastham, H. Hricak, J. A. Koutcher, and K. L. Zakian, "Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of ^{1}H magnetic resonance spectroscopic imaging," *J. Magn. Reson. Imag.*, vol. 40, no. 6, pp. 1414–1421, Dec. 2014.
[55] J. Wang, C.-J. Wu, M.-L. Bao, J. Zhang, X.-N. Wang, and Y.-D. Zhang, "Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer," *Eur. Radiol.*, vol. 27, no. 10, pp. 4082–4090, Oct. 2017.
[56] T. H. Nguyen, S. Sridharan, V. Macias, A. Kajdacsy-Balla, J. Melamed, M. N. Do, and G. Popescu, "Automatic Gleason grading of prostate cancer using quantitative phase imaging and machine learning," *J. Biomed. Opt.*, vol. 22, no. 3, Mar. 2017, Art. no. 036015.
[57] O. Hamzeh, A. Alkhateeb, J. Zheng, S. Kandalam, and L. Rueda, "Prediction of tumor location in prostate cancer tissue using a machine learning system on gene expression data," *BMC Bioinf.*, vol. 21, no. S2, pp. 1–10, Mar. 2020.

SAQIB IQBAL received the M.C.S. and M.Phil. degrees in computer science from the University of Azad Jammu and Kashmir Muzaffarabad, Pakistan, in 2013 and 2017, respectively. He is currently pursuing the Ph.D. degree in computer science with Quaid-I-Azam University, Islamabad, Pakistan. He taught computer science subjects as a visiting Junior Lecturer at the University of Azad Jammu and Kashmir. His main research areas are machine learning, biomedical image processing, data science, and virtual agents.

GHAZANFAR FAROOQ SIDDIQUI received the Ph.D. degree from Vrije Universiteit Amsterdam, in 2010. He is currently an Assistant Professor with the Department of Computer Science, Quaid-i-Azam University, Islamabad. Earlier, he was a Research Scholar with the Department of Computer Science, Vrije Universiteit Amsterdam, The Netherlands. The Ph.D. degree scholarship was funded by Higher Education Commission, Pakistan. He is a reviewer of a number of peer reviewed conferences and journals. He also published numerous research papers in reputed conferences and Journals. He is a member of Federal Public Service Commission and Khyber Pakhtunkhwa Public Service Commission. He is also serving as an evaluator for scientific projects of Directorate of Science and Technology (DoST) - Khyber Pakhtunkhwa. Furthermore, he is a member of Board of Studies of Quaid-i-Azam University, Islamabad, and National Textile University, Faisalabad.

AMJAD REHMAN (Senior Member, IEEE) received the Ph.D. degree from the Faculty of Computing, Universiti Teknologi Malaysia, in 2010, with a specialization in forensic documents analysis and security. He is currently a Senior Researcher with the AIDA Laboratory, Prince Sultan University, Riyadh, Saudi Arabia. He is the author of more than 100 indexed articles. His research interests are in data mining, health informatics, and pattern recognition.

LAL HUSSAIN received the M.S. degree (Hons.) in communication and networks from Iqra University, Islamabad, Pakistan, in 2012, and the Ph.D. degree from DCS & IT UAJK, in February 2016. He is an Assistant Professor with DCS & IT UAJK, Muzaffarabad, Pakistan. He worked as a Visiting Ph.D. Researcher at Lancaster University, U.K., for six months under HEC International Research Initiative Program and worked under the supervision of Dr. Aneta Stefanovska, a Professor of biomedical physics with the Physics Department, Lancaster University, U.K. He recently collaborated with Stony Brook University, USA, in different on-going projects with Prof. Tim. Q. Duong, since January 2020. He is the author of more than 30 publications of highly reputed peer-reviewed and Impact Fact Journals as a Principal author. He attended various conferences at tasks in ASIA (Pakistan) and Europe (USA and UK). He presented various talks in Pakistan, U.K., and USA.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently serving as an Associate Professor and the Associate Chair of the Information Systems Department, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia. She has more than 200 publications that have around 5000 plus citations with H-index 45. Her most publications are in the biomedical research published in ISI/SCIE indexed.

USMAN TARIQ received the Ph.D. degree in information and communication technology in computer science from Ajou University, South Korea. He is a skilled Research Engineer. He has a strong background in ad hoc networks and network communications. He is experienced in managing and developing projects from conception to completion. He have worked in large international scale and long-term projects with multinational organizations. He is currently attached with Prince Sattam Bin Abdul-Aziz University as an Associate Professor with the College of Computer Engineering and Science.

ADEEL AHMED ABBASI received the B.S.(Cs.) and M.Phil. degrees from the University of Azad Jammu and Kashmir Muzaffarabad. His main interests are machine learning, deep learning, and medical image processing.