KSÜ Mühendislik Bilimleri Dergisi, 27(4), 2024
Araştırma Makalesi

KSU J Eng Sci, 27(4), 2024
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 26.03.2024
Kabul Tarihi : 06.05.2024

Received Date : 26.03.2024
Accepted Date : 06.05.2024

DIAGNOSIS OF PROSTATE CANCER WITH ENHANCED EFFICIENCY USING FINE-TUNED CNN AND TRANSFER LEARNING

İNCE-AYAR İLE ETKİNLİĞİ ARTIRILMIŞ ESA VE TRANSFER ÖĞRENME YÖNTEMLERİYLE PROSTAT KANSERİNİN TESPİTİ

Murat SARIATEŞ[1] (ORCID: 0009-0003-4351-9566)
Erdal ÖZBAY[1*] (ORCID: 0000-0002-9004-4802)

[1] Fırat Üniversitesi, Bilgisayar Mühendisliği Bölümü, Elazığ, Türkiye

*Sorumlu Yazar / Corresponding Author: Erdal ÖZBAY, erdalozbay@firat.edu.tr

ABSTRACT

Cancer is one of the high-risk diseases for humans. Prostate cases are the second most common disease in men after lung cancer, and early diagnosis is vital. Artificial intelligence technologies have begun to be used in the diagnosis of prostate cancer, and more effective and sensitive results have been obtained, preventing potential errors in human-centered methods. In this study, in order to increase the classification performance in the diagnosis of prostate cancer, transfer learning methods and fine-tuning processes, which have higher success and learning ability with less training data, unlike machine learning methods, were applied. The two-class data set consisting of prostate cancer MR images, 'significant' and 'not-significant', was classified with Alexnet, Densenet201, Googlenet, and Vgg16 models with the feature extraction approach, and 71.40%, 72.05%, 65%, and 80.13% accuracy results were obtained respectively. To increase these rates, pre-trained transfer learning models were used and accuracy results of 89.74%, 94.32%, 85.59%, and 91.05% were achieved, respectively. A 98.10% validation result was obtained using the cross-validation method in the Densenet201 model. DenseNet201 model achieved the highest accuracy result of 98.63% in transfer learning with the combination of the RMSProp optimization method. The proposed transfer learning model provided an improvement of approximately 26% compared to the feature extraction method.

Keywords: Prostate cancer, CNN, deep learning, classification, transfer learning

ÖZET

Kanser insanlar için yüksek riskli hastalıkların başındadır. Prostat vakaları, akciğer kanserinden sonra erkeklerde ikinci sırada yer almakta ve erken teşhisi hayati önem taşımaktadır. Prostat kanserinin teşhisinde yapay zeka teknolojilerinden faydalanılmaya başlanmış, daha etkili ve hassas sonuçlar elde edilerek insan odaklı yöntemlerdeki potansiyel hatalarının önüne geçilmiştir. Bu çalışmada prostat kanserinin teşhisinde sınıflandırma performansını arttırabilmek adına makine öğrenmesi yöntemlerinden farklı olarak daha az eğitim verisi ile daha yüksek başarı ve öğrenme kabiliyetine sahip transfer öğrenme yöntemi ve ince-ayar işlemleri uygulanmıştır. Prostat kanseri MR görüntülerinden oluşan 'significant' ve 'not-significant' olmak üzere iki sınıflı veri setine, özellik çıkarımı yaklaşımıyla Alexnet, Densenet201, Googlenet ve Vgg16 modelleriyle sınıflandırılarak sırasıyla %71,40, %72,05, %65,72 ve %80,13 doğruluk sonuçları elde edilmiştir. Bu oranları arttırabilmek adına ön-eğitimli transfer öğrenme modelleri kullanılmış ve sırasıyla %89,74, %94,32, %85,59 ve %91,05 doğruluk sonuçlarına ulaşılmıştır. Densenet201 modelinde çapraz-doğrulama yöntemi kullanılarak %98,10 doğrulama sonucu elde edilmiştir. DenseNet201 modeli transfer öğrenmede RMSProp optimizasyon yöntemi kombinasyonuyla %98,63 ile en yüksek doğruluk sonucuna ulaşmıştır. Önerilen transfer öğrenme modeli, özellik çıkarımı yöntemine kıyasla yaklaşık %26 oranında bir iyileştirme sağlamıştır.

Anahtar Kelimeler: Prostat kanseri, ESA, derin öğrenme, sınıflandırma, transfer öğrenme

ToCite: SARIATEŞ, M., & ÖZBAY, E., (2024). DIAGNOSIS OF PROSTATE CANCER WITH ENHANCED EFFICIENCY USING FINE-TUNED CNN AND TRANSFER LEARNING. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 27(4), 1306-1319.

INTRODUCTION

Worldwide and in our country, cancer is the second leading cause of death after cardiovascular diseases with a prevalence of 22%. Although the incidence of cancer varies by gender, it tends to be 25% more common in men than in women (Dorak and Karpuzoglu, 2012).

Prostate cancer is also particularly common in men and occurs as a result of abnormal growth and proliferation of cells in the prostate gland. While normal prostate cells grow and divide as much as the body needs, cancerous cells lose this control and begin to multiply rapidly and abnormally. In this way, prostate cancer can spread to other organs or lymph nodes outside the pelvic region and as a result of this metastasis, it causes an increase in Prostate Specific Antigen (PSA) levels in the blood (Carlsson et al., 2014).

PSA tests contribute to the early diagnosis of prostate cancer but can only be elevated in many benign conditions such as Benign Prostatic Hyperplasia. Rectal examination can detect tumors above 0.2 ml but has low sensitivity. In recent years, Multiparametric Prostate Magnetic Resonance Imaging (MpMRI) techniques have become more important in the diagnosis and treatment of prostate cancer (Bjurlin et al., 2020). In addition to imaging techniques, treatment options include surgical intervention, radiation therapy, hormone therapy, chemotherapy, and immunotherapy. The choice of treatment usually depends on the stage of the cancer, the general health status of the patient, and other factors.

When prostate cancer is diagnosed at an early age, the survival rate after treatment increases significantly, but during the prolonged survival period, patients may also face various psychosocial problems such as stress, anxiety, depression, and social isolation due to the side effects of cancer treatment. In addition, physiological problems such as weight loss, fatigue, anorexia, and sleep problems may trigger psychosocial problems (Himmerich et al., 2021). Therefore, early detection of prostate cancer prolongs the patient's life expectancy and reduces the risk of possible complications by increasing treatment success.

This study is planned to effectively detect prostate cancer with CNN and transfer learning methods, which have increased efficiency with fine-tuning. In the intermediate stages of the study, statistical data such as sensitivity, AUC, F1-score, and ROC curves will be included to understand the difference in classification accuracy rates from other studies and the necessary analyses will be made in the light of this information.

To date, many classification processes have been performed using traditional machine learning techniques. Some commonly used algorithms in the field of machine learning are given below.

- Naive Bayes Classifier
- K-nearest neighbor algorithm
- Decision Trees
- Support Vector Machine
- Random Forest
- Linear Regression
- Logistic Regression
- K-means algorithm

Many studies have been conducted on prostate cancer prediction using KNN, SVM, LR, NB, and RF algorithms, and as a result of these studies, the accuracy results were found between 70% and 90%. The highest accuracy of 90% was obtained from RF and LR algorithms (Srivenkatesh, 2020).

When the necessary literature review is made, it is observed that the success rate increases when deep learning methods are used instead of traditional machine learning techniques.

One way to use the existing model faster and more efficiently is to use transfer learning techniques. This allows for generalization, fine-tuning, learning speed, and efficiency (Weiss et al., 2016). The difference between traditional machine learning and transfer learning is shown graphically in Figure 1.

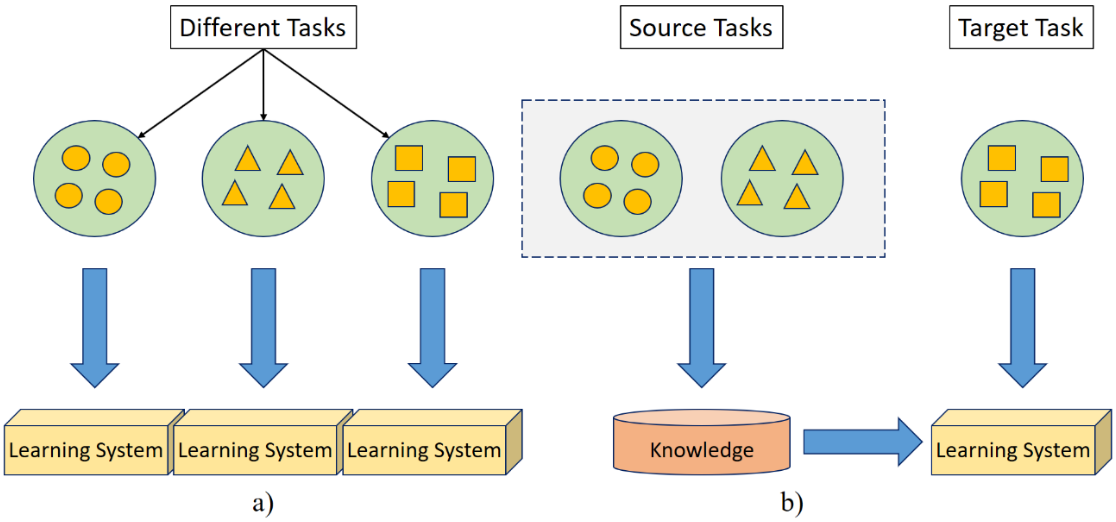

Figure 1. The Difference Between a) Traditional Machine Learning and b) Transfer Learning

When working with a non-large dataset, it has often made more sense to opt for a transfer learning approach to improve the performance of a deep learning model. Transfer learning is an approach that utilizes existing knowledge to improve the performance efficiency of limited learning data on the target task. This approach allows the knowledge already learned in the source task to be adapted to the training of the model in the target task, making it possible to achieve better results despite the limited amount of data. Therefore, the best trend of deep learning is to make the best use of the available knowledge by using transfer learning in limited data situations (Zhuang et al., 2020).

The approach realized in this study is aimed to diagnose prostate cancer faster from MR images and to allow more time to perform the necessary intervention. In addition, it is aimed to diagnose cancer at an early stage and to prevent the patient from being psychologically worn out by other methods applied during the diagnosis phase. The article's contributions to the literature can be summarized as follows:

• Classification performances obtained by hand-crafted feature extraction of pre-trained CNN architectures in prostate cancer diagnosis have been demonstrated.
• Transfer learning methods and fine-tuning processes have been applied with high classification success with insufficient or small training datasets.
• The DenseNet201 model achieved the highest accuracy result of 98.63% in transfer learning with the combination of the RMSProp optimization method using 5-fold CV.
• The proposed transfer learning method with fine-tuning efficiency, achieves much superior performance than hand-crafted feature extraction approach CNN models in prostate cancer diagnosis.

The study aims to contribute to other disease classification methods in the literature and to be useful for new studies. We focus on previous research to understand the context and significance of the current study, summarize previous work on our topic, and provide a better understanding of what gaps the current research fills, what questions it answers, and what aspects it contributes to.

A review of the literature reveals that disease detection and similar studies have generally been conducted using traditional machine learning methods, while transfer learning models have been popularly used in these classification and detection methods, especially in the last five years.

Swati et al. studied brain tumor classification for MRI images using transfer learning and fine-tuning. In this study, a pre-trained deep CNN model was used and a block-wise fine-tuning strategy based on transfer learning was proposed. An average accuracy of 94.82% was obtained (Swati et al., 2019).

Aslan et al. studied deep learning-based automatic brain tumor classification and obtained an accuracy of 96.44% in the classifier by using the MobilNetV2 deep learning model and k-nearest run (k-EYK) algorithm (Aslan, 2022).

Kiliçarslan et al. used transfer learning methods for disease detection in tomato leaves and the highest result of 99% was obtained with DenseNet architecture in studies conducted with DenseNet, ResNet50, and MobileNet architectures (Kılıçarslan and Pacal, 2023).

All these studies show that transfer learning methods can be used effectively in areas such as disease detection and classification. The current study was prepared by reviewing the necessary literature and examining similar studies.

MATERIAL AND METHOD

In this study, an automatic diagnosing system was designed with pre-trained CNN architectures and fine-tuning transfer learning methods to effectively detect prostate cancer. The flow diagram of the proposed method is illustrated in Figure 2.

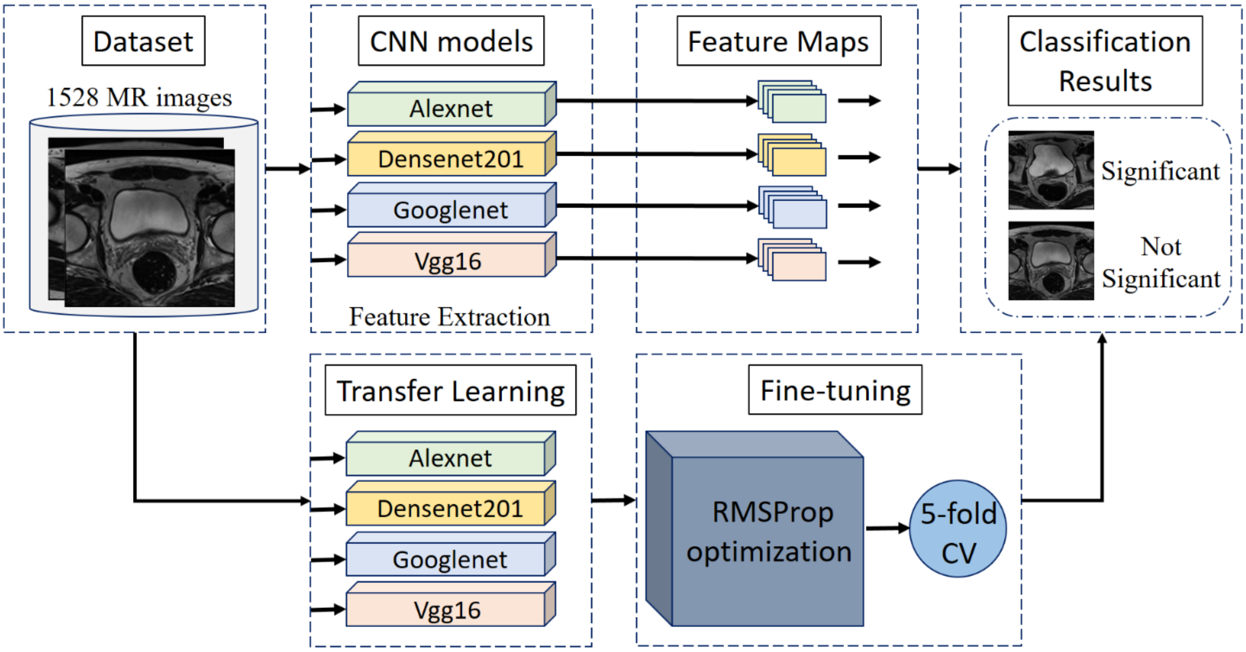

Figure 2. The Flow Diagram of the Proposed Methodology

Dataset

In this study, prostate cancer classification processes were performed using transfer learning with increased efficiency CNN, which has become increasingly popular in the literature. The dataset used to obtain the experimental results in this study is MRI images of prostate cancer taken from the Kaggle website (Geert et al., 2017). Dataset is divided into two classes as 'significant' and 'not significant'. This dataset consists of 764 'significant' images, i.e. clinically significant, and 764 'not significant' images, i.e. clinically insignificant, totaling 1528 MR images. The image data, which was initially received as 8 bits, was set to 24 bits and used for classification. A sample of two different classes of the dataset is shown in Figure 3.

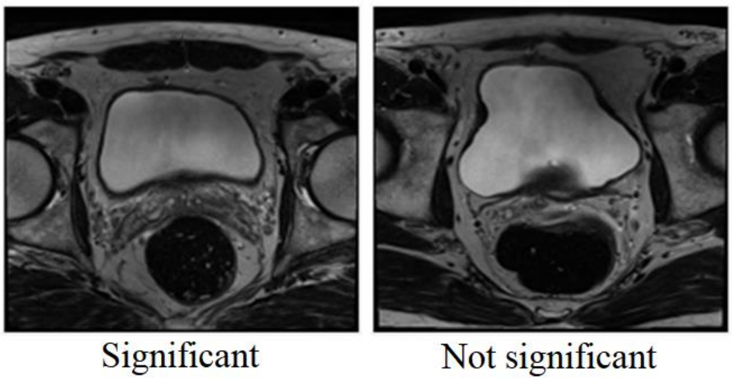

Figure 3. Sample MRI Images of the Prostate Dataset

Convolutional Neural Networks (CNN)

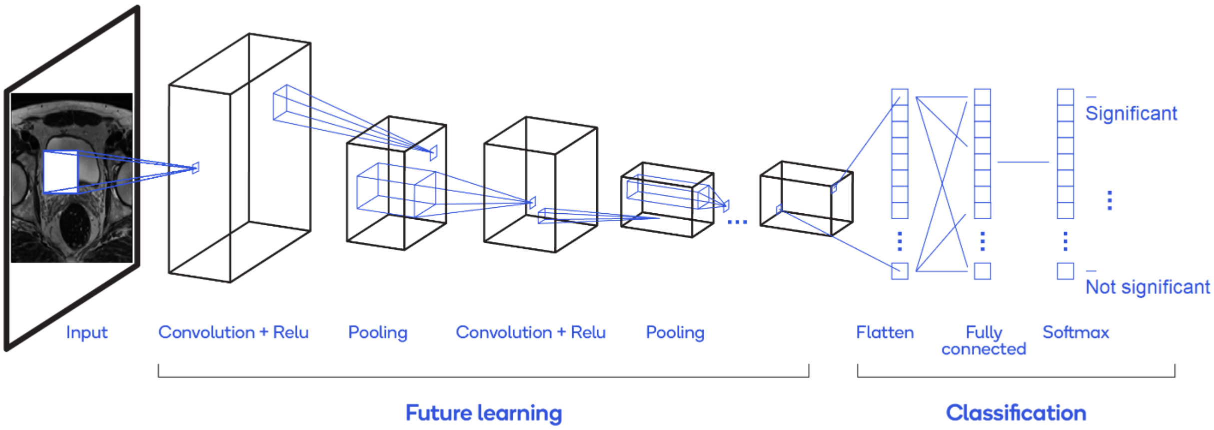

Figure 4. CNN Architecture and Layers

Convolutional Neural Networks (CNNs) are deep learning architectures based on artificial neural networks. CNN architectures contain more layers and neural cells than neural networks and therefore require a higher computational cost to train. However, CNNs have proven classification success. A common method is to transfer weights from the layers of a CNN trained with a large and comprehensive training dataset. This is the process of transferring the learned features of a pre-trained CNN model for a specific task to another task (Fırıldak and Talu, 2019). The classical CNN architecture and layers are shown in Figure 4.

The basic layer types used in CNN are as follows (Zainudin et al., 2020):
• Convolutional Layer: It is the layer that learns the features in different parts of the image by shifting a window (filter or kernel) of a certain size on the image.
• ReLU Layer (Rectified Linear Unit Layer): This layer uses the ReLU function as the activation function. This layer helps CNN to learn non-linear features.
• Pooling Layer: it is used to reduce the size of the feature maps produced by convolutional layers.
• Fully Connected Layer: Used to convert the output of the CNN into an output for classification or regression. They are usually the last layers and enable the network to learn.
• Dropout Layer: The dropout layer is used to prevent overfitting of the network.
• Normalization Layers: used to increase the stability of the network and speed up the training process.

Classification with CNN

Image classification is a machine learning and artificial intelligence technique used to identify and assign the content of a digital image to a specific category or class. This is accomplished using a computer algorithm that automatically analyzes and recognizes the content of an image.

Table 1. The Parameters of the CNN Models

Parameters	**Values**
Batch size	10
Max epoch	6
Learning rate	0.0001
Validation frequency	3
Optimization	SGDM

In this study, to better understand the performance of the trained model developed with transfer learning, classification was first applied to the existing dataset. Of the two-class data set, 70% is allocated for training and 30% for testing. In the architecture, the fully connected layer is set appropriately and the classification layer is redesigned according to the input data. To train a network using SGDM (Stochastic Gradient Descent with Momentum), the size of the minibatch to be used for each training iteration is set to 'minibatchsize=10', the maximum number of epochs to be used for training is set to 'maxepochs=6', the initial learning rate used for training is set to 'initiallearnrate=0.0001', the neural network validation frequency in terms of the number of iterations is set to ‘validationfrequency=3’. The parameters of the CNN model are given in Table 1. These settings were the same for Alexnet, Densenet201, Googlenet, and Vgg16 architectures and the classifications were made. The dataset of prostate cancer MRI images was classified by making the necessary adjustments in Alexnet, Densenet201, Googlenet, and Vgg16 architectures by feature extraction method.

Transfer Learning and Fine-tuning

Transfer learning, as a machine learning concept, generally involves using knowledge gained in one task as the starting point of a model in another task. For example, if an image recognition model has been trained on the AlexNet dataset and has gained general object recognition capabilities, its feature extraction layers or pre-trained weights can be used as a starting point for a different task. This can lead to better results using less training data for the new task.

Transfer learning has been very useful when the dataset size is small, when there is not enough training data for the new task, or when it is necessary to use pre-trained models for a specific task. Moreover, transfer learning can improve the generalization ability and help the model perform better on real-world data.

In this study, Alexnet, Densenet201, Googlenet, and Vgg16 architectures were used for transfer learning. In these models, the dataset is split for training and validation, with 70% of the data used for training and 30% for validation. For example, the last three layers of the AlexNet model were removed and replaced with a new fully connected layer, a softmax layer, and a classification layer. This customized the model for the classification of prostate images.

Data augmentation was performed by augmenting the images in a predetermined way with techniques such as random rotation, panning, and mirroring. This helps the model learn more general features and reduces overfitting.

Training options were identified. These options determine which optimization algorithm is used during training, the learning rate, how often the validation data is evaluated, and how many epochs the training continues.

Optimization algorithms try to improve the training process of the model by managing the learning process (Özbay 2023). The correct choice of these algorithms can affect the speed and stability of the training process and the quality of the results. The correct choice of the learning coefficient is critical for training the model quickly and at the same time achieving the minimum point. A small learning coefficient can slow down the solution process and increase the training time. A large learning coefficient can lead to skipping the minimum point and excessive fluctuations (Seyyarer et al., 2020).

In this study, SGDM (Stochastic Gradient Descent Moment) was initially used as the most suitable optimization algorithm. However, it was later observed that higher accuracy rates were obtained when RMSProp (Root Mean Square Propagation) was selected as the optimization algorithm. RMSProp (Root Mean Square Propagation) is a gradient-based optimization algorithm widely used in deep learning and optimization algorithms. RMSProp can be considered an extension of the widely used Stochastic Gradient Descent (SGD) algorithm. The main goal of RMSProp is to correct the tendency of SGD to behave slowly at low dimensional learning rates and unstable at large dimensional learning rates. This is done by calculating the exponential moving average of the gradient squares. This method normalizes the magnitude of the gradients by adding a scaling term in which the gradients are divided by the average of their previous squares. Research has also shown that the optimal value for the initial learning rate is 0.0001. With necessary fine-tuning, the size of the minibatch was set to ‘minibatchsize=10’, the maximum number of epochs to be used for training was set to ‘maxepochs=6’, and the neural network validation frequency in terms of the number of iterations was set to ‘validationfrequency=3’. The parameters of the Transfer learning models are given in Table 2.

Table 2. The Parameters of the Transfer Learning Models

Parameters	**Values**
Batch size	10
Max epoch	6
Learning rate	0.0001
Validation frequency	3
Optimization	RMSProp

Cross-Validation

An important method for reliably evaluating classification models built with transfer learning is the k-fold cross-validation method. Cross-validation is a method used to evaluate the performance of machine learning models. This method is used to verify the generalization ability of the model and helps to prevent overfitting.

Overfitting is when a machine learning model fits the training data too well and overly specializes in the training data (Montesinos et al., 2022). This reduces the generalization ability of the model and can lead to poor performance on new, unseen data. Figure 5 shows how the classification decision boundary follows the training data too closely for an overfitting model and not closely enough for an underfitting model.

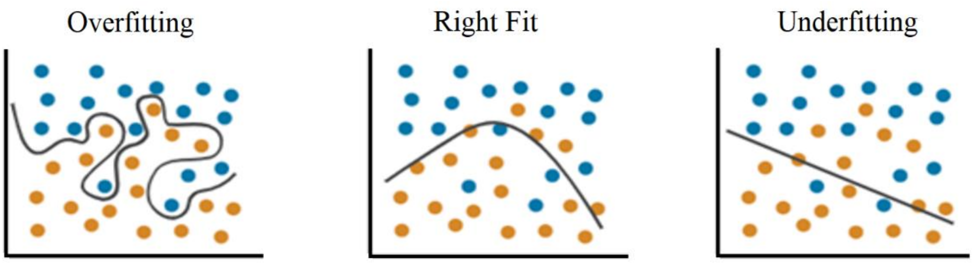

Figure 5. Overfitted Models of Classification

Training a model on a single dataset and evaluating its performance on the same dataset may not provide a reliable estimate of its generalization ability. k-fold cross-validation overcomes this challenge by splitting the dataset into k-folds and ensuring that each fold maintains the same class distribution as the original dataset. This technique allows for a more robust evaluation of models by training and testing models on multiple subsets of data (Mahesh et al., 2023).

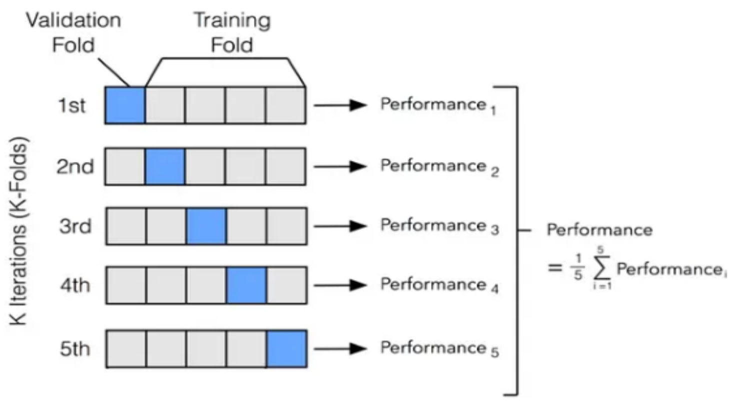

Figure 6. K-fold Cross-Validation [16]

In this study, the k-fold value was set to 5. It is an important point which part of the existing data set is taken as training data and which part is taken as test data. For this reason, in order not to cause any error and to reach the real accuracy value, a 5-step test, and training data were determined for the dataset on the trained models. In this way, the overfitting problem of the model was tried to be solved (Anguita et al., 2012). The k-fold cross-validation working methodology is shown in Figure 6.

EXPERIMENTAL RESULTS AND DISCUSSION

In this study, the classification algorithms described in section 2 were applied to the dataset described in section 3 for the detection of prostate cancer. Before looking at the outputs of the results obtained in transfer learning, the results obtained with the image classification method with Alexnet, Densenet201, Googlenet, and Vgg16 architectures are evaluated.

Performance Metrices

Initially, image classification was performed for Alexnet, Densenet201, Googlenet, and Vgg16 architectures, and then with a transfer learning model. Experiments were conducted to see the effect of the selected parameters and the trained model on the classification accuracy and the results are reported. To compare the results and make better choices, accuracy metrics are used as model evaluation metrics. Accuracy is a proportional measure of the agreement between the predicted class and the true class. For example, accuracy is presented as the ratio of the number of cases predicted by the model to be at 'significant' risk for prostate cancer to the proportion of all cases that are actually at risk and the proportion of cases predicted by the model to be at no risk that is actually at no risk. Classification accuracy is derived from the metrics specified in the complexity matrix shown in Table 3.

Table 3. Two-class Confusion Matrix

		Predicted Class	
		Risk exist	No risk
True class	Risk exist	TP (True Positive)	FN (False Negative)
	No risk	FP (False Positive)	TN (True Negative)

Complexity matrix terms and their meanings are listed below:
- True Positive (TP): Represents the amount of data that belongs to the positive class and is correctly classified by the classifier.
- True Negative (TN): Represents the amount of data belonging to the negative class that was correctly classified by the classifier.
- False Positive (FP): An expression that actually belongs to the negative class is misclassified as a positive class.
- False Negative (FN): This is the misclassification of a statement as a negative class when it actually belongs to the positive class.

The metrics determined using the complexity matrix to determine the classification performance are listed below with their explanations (Özbay & Özbay, 2021):
Accuracy: It is used to evaluate the success of the proposed model. An accuracy value is calculated by dividing the set of correct predictions in the model by the entire dataset. It is calculated using performance evaluation metrics as given in Equation 1.

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

Precision: Especially important when the cost of false positive predictions is high. For example, if the model marks as spam (FP) an email that should arrive in your inbox, you will not see the important email you should receive and you will be at a loss. Precision is an important criterion when choosing a model. It is calculated as given in Equation 2.

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

Recall: A useful metric also when the cost of predicting false negatives is high. It should be as high as possible. It is calculated as shown in Equation 3.

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

F-score: This value is the harmonic mean of the sensitivity and precision (recall) values. It is calculated as given in Equation 4.

$$F1\ score = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{4}$$

ROC curve (Receiver Operating Characteristic curve): a graphical tool used to evaluate the performance of a model in classification problems. In particular, it is used to visualize the sensitivity and specificity of the model. The ROC curve shows the relationship between sensitivity and specificity at different cut-off points (thresholds). The X-axis shows the sensitivity, defined as the false positive rate (FPR) and the Y-axis shows the true positive rate (TPR).

AUC (Area Under the ROC Curve): Refers to the area under the ROC curve. This value is a measure of the performance of classifiers. The AUC value can be between 0 and 1. If the AUC value of a model is close to 0.5, it can be said that it is as if it is performing random classification. However, as the AUC value approaches 1, the performance of the model improves.

Performance Evaluation

When calculating the classification accuracy, it is important to determine its effectiveness with the tests performed. In this respect, while determining the accuracy, the performance of the model is also examined. The complexity matrix was used to determine this performance. Then, using this matrix, accuracy, precision, sensitivity, and F-score metrics are calculated. The classification process was performed by making the necessary adjustments in Alexnet, Densenet201, Googlenet, and Vgg16 architectures with the feature extraction method, and the validation results were obtained as 71.57%, 72.05%, 65.72%, and 80.13% respectively. Confusion matrices of experimental results are given in Figure 7. The highest accuracy rate was obtained from the Vgg16 architecture with 80.13%.

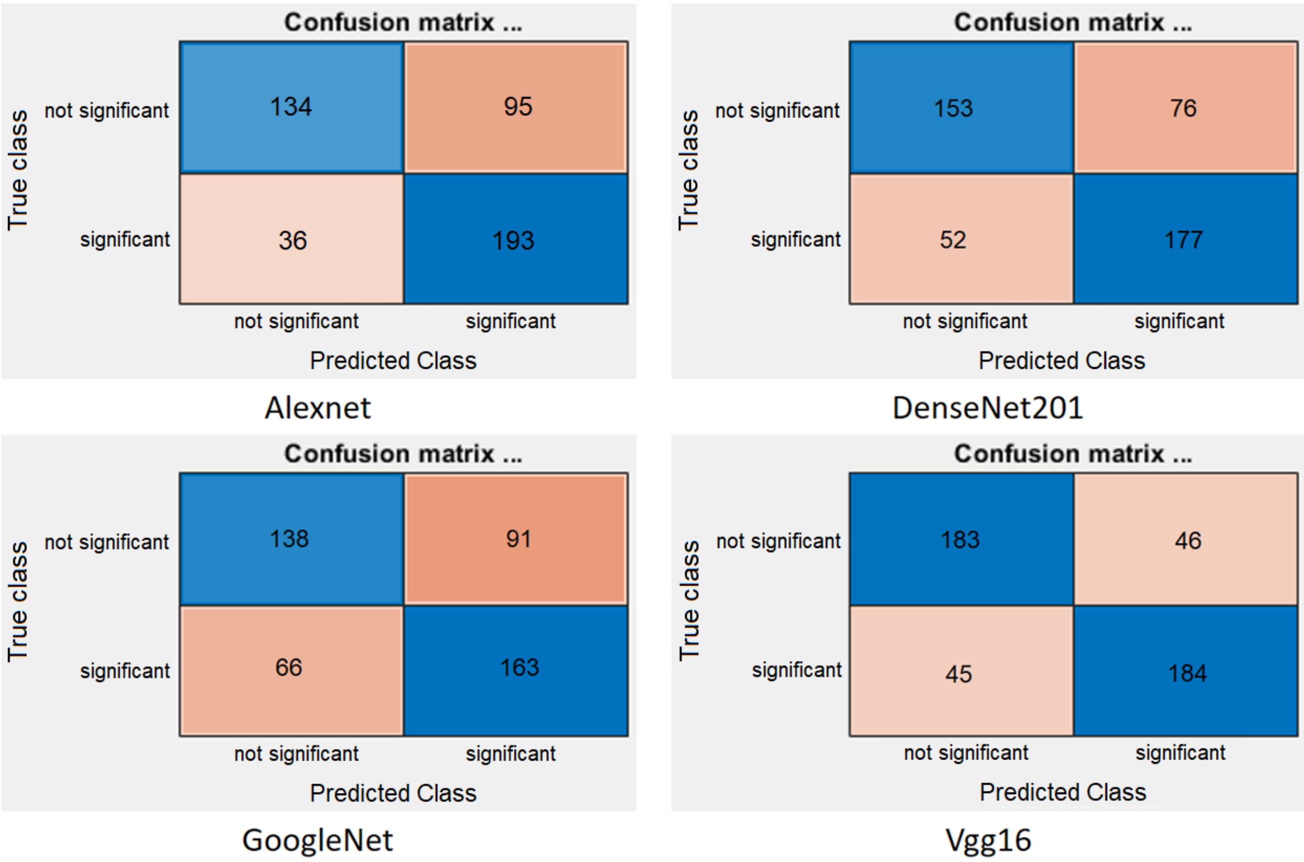

Figure 7. Confusion Matrices of the Pre-trained Models

The table of performance metrics calculated for Alexnet, Densenet201, Googlenet, and Vgg16 architectures is given in Table 4.

Table 4. Results Obtained Through Evaluation Metrics

Models	**Accuracy**	**Precision**	**Recall**	**F1-score**
AlexNet	71.40%	58.52%	78.82%	67.17%
DenseNet201	72.05%	66.81%	74.63%	70.51%
GoogleNet	65.72%	60.26%	67.65%	63.74%
Vgg16	80.13%	79.91%	80.26%	80.09%

When the performance metrics mentioned above are analyzed in Table 4, it is seen that the highest classification accuracy is obtained with the Vgg16 architecture. To increase these validation results, transfer learning methods were applied and classification processes were performed again on the same architectures with the trained models. The models were trained with the transfer learning method and then the accuracy results were obtained for Alexnet, Densenet201, Googlenet, and Vgg16 architectures given in Table 5.

Table 5. Accuracy Results of the Transfer Learning Methods

Models	AlexNet	DenseNet201	GoogleNet	Vgg16
Accuracy	89.74%	94.32%	85.59%	91.05%

Initially, 70% of training data and 30% of test data are taken, so it is not clear exactly which part of the dataset is training data and which part is test data. For this reason, a machine learning model can adapt too much to the training data and become overly specialized to the training data. In order to avoid this situation and to get more accurate results, the k-fold cross-validation method was used. With this method, the data set for transfer learning models is divided into 5 parts, and each part is used as test and training data respectively. The performance metrics were obtained for Alexnet, Densenet201, Googlenet, and Vgg16 architectures given in Table 6 with this method.

Table 6. Performance Results of Transfer Learning Models with K-fold Cross-Validation

Models	Accuracy	Precision	Recall	F1-score
AlexNet	88.22%	85.87%	91.49%	88.59%
DenseNet201	98.10%	99.60%	96.60%	98.07%
GoogleNet	84.23%	84.73%	83.51%	84.11%
Vgg16	98.30%	98.30%	98.30%	98.30%

As seen in Table 6, the highest accuracy rate was obtained with Vgg16 at 98.30%. Accordingly, the ROC curve of the result obtained with Vgg16 is given in Figure 8.

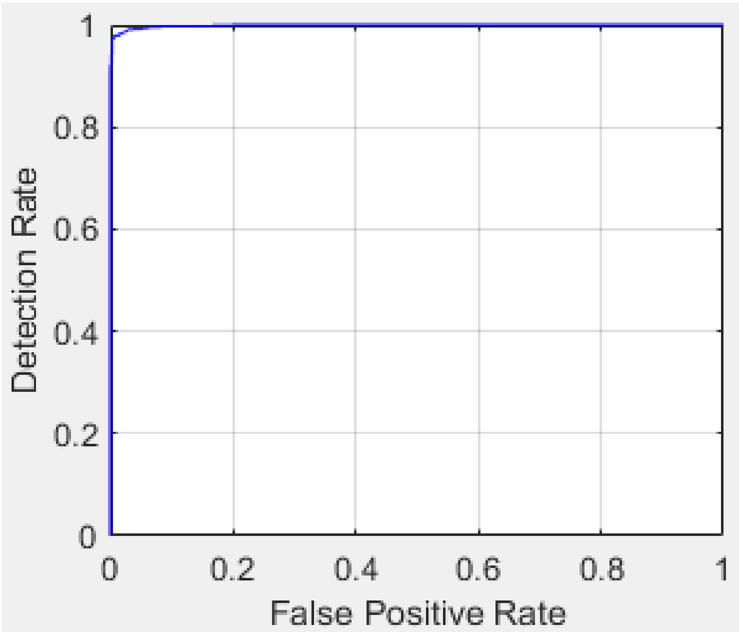

Figure 8. ROC Curve of Vgg16 Using Transfer Learning with K-fold Cross-Validation

The shape of the ROC plot provides important clues for interpreting the performance of the model. For an ideal model, the ROC curve passes through the upper left corner, meaning that the true positive rate is high and the false positive rate is low. The closer the ROC curve is to this ideal, the better the performance of the model.

Accordingly, the SGDM used as the optimization algorithm was modified and the RMSProp optimization algorithm was tested by making appropriate adjustments for the DenseNet201 and Vgg16 architectures with the two highest accuracy rates. In the examinations, higher accuracy results were obtained with the selected optimization algorithm. The transfer learning performance of the DenseNet201 model was obtained using the RMSProp optimization algorithm. Figure 9 shows the confusion matrix and ROC curve of DenseNet201's transfer learning performance with RMSProp.

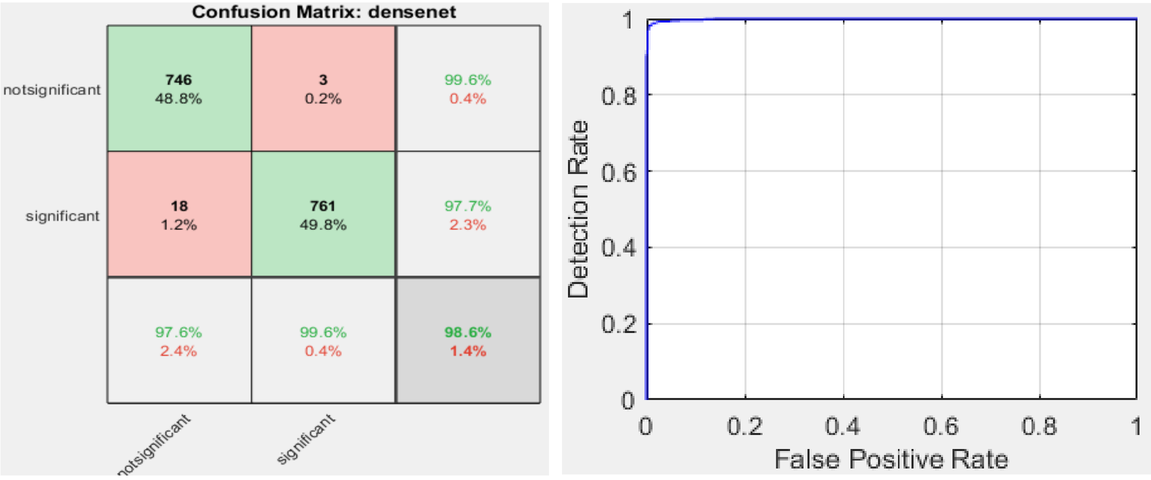

Figure 9. Confusion Matrix and ROC Curve of Densenet201 Using RMSProp Optimization

When the results given in Figure 9 were analyzed, it was seen that the highest accuracy rate of 98.63% was achieved with DenseNet201. When these results are evaluated in terms of performance metrics, the DenseNet201 model achieved the results in Table 7 by using the RMSProp optimization method in transfer learning.

Table 7. Performance Results of Densenet201 Optimized with RMSProp

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
DenseNet201	98.63%	99.60%	97.64%	98.61%

According to the evaluation results given in Table 7, the DenseNet201 model used the RMSProp optimization method in transfer learning achieved the result with the highest accuracy, precision, recall, and F1-score of 98.63%, 99.60%, 97.64%, and 98.61%, respectively. In this regard, an improvement of approximately 26% was achieved in the classification accuracy rate achieved with the proposed transfer learning model compared to the classification accuracy rate obtained with the feature extraction method.

The use of transfer learning and deep learning models in prostate cancer classification has an important place in the current literature. When the necessary research was done, it was seen that transfer learning methods were used in prostate cancer detection. However, when the contributions of this study are examined; In terms of optimization strategies, using optimization algorithms such as RMSProp as well as SGD and comparing the performance of these algorithms have been positive to obtain accurate results. Dividing the data set with k-fold cross-validation enabled a more reliable evaluation of the model’s performance, and also helped the model to be more resistant to overfitting and increased its generalization ability. The high accuracy rates obtained especially when the RMSProp optimization algorithm was used with the Densenet201 architecture showed the effectiveness and potential of this method. A total of 1528 data sets used in the study, divided into two classes as ‘significant’ and ‘not significant’, were evaluated to be successful due to the high accuracy rate of the transfer learning method used, although their number was lower than the data sets of other studies. The 98.63% accuracy rate obtained in this study, especially when the Densenet201 architecture and RMSProp optimization algorithm were used, showed that the current study was successful compared to other studies. In addition, the proposed method is compared with other state-of-the-art studies in the literature in Table 8.

The proposed method was compared with similar approaches to classify prostate cancer diseases, some of which are described in the literature section. In the proposed method, the features of the input image can be learned with different pre-trained CNN architectures. The feature obtained by feature extraction methods was classified with both CNN and transfer learning algorithms. Additionally, the cross-validation and RMSProp optimization algorithm was used to optimize the CNN parameters on transfer learning, which increased model robustness and accelerated convergence. The methods of the existing studies from recent years used the same or similar data sets. In this respect, when compared to the recent state-of-the-art studies given in Table 8, it can be said that the proposed method has greater potential than existing approaches due to its ease of application compared to traditional methods, its ability to handle multi-class variance, and its high classification rate.

Table 8. Performance Results of the Proposed Method with Current State-of-the-Art Studies

Authors / Year	Method	Result
(Abdelmaksoud et al., 2021)	Improved VGGNet	91.20%
(Yuan et al., 2019)	MPTL	86.92%
(Zhong et al., 2019)	DTL-based model	89.00%
(Kanna et al., 2023)	DL-based models	84.99%
(Tsuneki et al., 2022)	EfficientNetB1	97.80%
(Abbasi et al., 2020)	GoogleNet, SVM, SIFT	99.71%
(Hoar et al., 2021)	Boost CNN	93.00%
(Chavda and Degadwala, 2024)	ResNet and VGG	89.00%
(Hamm et al., 2023)	XAI, PI-RADS	80.00%
(Proposed method, 2024)	Transfer Learning with RMSProp	98,63%

Training data for deep learning architectures must be abundant. Furthermore, labeled data is needed for supervised learning in the majority of deep learning approaches, such as CNN-based methods, which is challenging and time-consuming in the clinical setting. It is still unclear how to efficiently train deeper networks and make the most of the training set's small amount of data. The literature has two popular solutions that can address the aforementioned issue in part. The first is data augmentation, which creates new data from the existing data by using affine transformations including translation, rotation, and scaling. However, in this study, we present an approach that can produce effective results on restricted data. The other approach is transfer learning, which has shown promise in the processing of medical images. There are two components to the transfer learning workflow: adjusting using the intended dataset and pre-training such as DenseNet201.

This study provides a brief evaluation of the future scope of prostate cancer diagnosis. Even with the difficulties CAD systems in clinical settings provide and the advent of deep learning techniques, the encouraging outcomes are too valuable to ignore. By extracting knowledge from large amounts of data, deep learning techniques generate an output that may be utilized for individualized treatment, thereby advancing precision medicine. Unlike traditional medical care, precision medicine focuses on the tiniest molecular and genomic details, and medical professionals base their diagnosis decisions on minor variations between patients. Radiomics emerged with the advancement of big data and medical imaging. With the use of several medical pictures and feature-related algorithms, it seeks to convert the region of interest into high-resolution feature maps. In the future, medical images referred to as imaging grouping will be easily linked to non-imaging data in electronic medical records, such as gender, age, medical history, and so forth. When applied to electronic medical data, deep learning techniques can help derive patient representations that could result in forecasts and enhancements of clinical decision support systems. Opportunities for wider use of CNN-based CAD systems in clinical practice exist because of the recent and rapid development of deep learning technology, particularly CNN-based approaches. These methods are not anticipated to replace radiologists in the near future, but they may ease normal workflow, increase the precision of diagnosis and detection, lower the likelihood of errors, and improve patient satisfaction.

For the sake of a fair evaluation, it was considered that it would be useful to address the weaknesses of the proposed approach. Accordingly, in the current study, no detailed analysis was made as to which features or image regions affect the classification performance of the model during the classification process. This type of analysis can help us better understand the impact of the model on the results. In addition, the inadequacy of the data set used and the image quality resulted in limited results. It was evaluated that increasing the mentioned features of the data set could allow positive results to be obtained on the results. When the results of the study are examined; It has been shown that high accuracy rates can be achieved with a limited data set. In this context, it is envisaged that new transfer learning methods to be developed in future studies will make a greater contribution to disease classification processes, shorten detection and diagnosis times, and more accurate results can be obtained regardless of the number of data sets.

Transfer learning may not provide significant benefits if the tasks in problem-solving are very different. If the source dataset is too small or not representative of the target domain, the transferred information may not be sufficient. If there is a significant difference between the distribution of the data between the source field and the target field, it may not be able to generalize well. To alleviate this problem, domain adaptation techniques are often used. Additionally, fine-tuning may require adjustments to the architecture, which can be complex and time-consuming, leading to overfitting, especially if the model has a large number of parameters. Despite these limitations, transfer learning remains a valuable tool in machine learning, particularly in scenarios where labeled data is limited or expensive to acquire.

CONCLUSIONS

In this study, we aimed to improve classification accuracy by using models trained with transfer learning on image data consisting of prostate cancer MRI dataset. Initially, without the use of trained models, feature extraction was performed on the dataset using Alexnet, Densenet201, Googlenet, and Vgg16 architectures. These results were obtained as 71.57%, 72.05%, 65.72%, and 80.13% respectively. Then, to compare the results, pre-trained models and the SGDM optimization algorithm were used together and necessary fine-tuning was done. With this method, 89.74%, 94.32%, 85.59%, and 91.05% were obtained for Alexnet, Densenet201, Googlenet, and Vgg16 architectures, respectively. Considering the overfitting on the dataset, 5-step dataset segmentation with the k-fold cross-validation method was applied to the transfer learning models in order to obtain more accurate rates. The accuracy rates obtained with this method were 88.22%, 98.10%, 98.10%, 84.23%, and 98.30% respectively. Considering the tendency of SGDM, which is used as an optimization algorithm, to behave slowly at low dimensional learning rates and unstable at high dimensional learning rates, RMSProp was chosen as the optimization method for the model trained with the DenseNet201 architecture to obtain the highest accuracy rate. Compared to other trained models, the highest accuracy rate was obtained with the DenseNet201 architecture at 98.63%. A performance increase of approximately 26% was achieved with the trained model. In this study, the classification of prostate cancer by transfer learning with CNN, whose efficiency was increased with fine-tuning, was carried out. We investigated how pre-trained deep learning models can be used in important applications such as disease classification. The results show that transfer learning has the potential to significantly improve classification performance.

REFERENCES

Abbasi, A. A., Hussain, L., Awan, I. A., Abbasi, I., Majid, A., Nadeem, M. S. A., Chaudhary, Q. A. (2020). Detecting prostate cancer using deep learning convolution neural network with transfer learning approach. *Cognitive Neurodynamics*, 14, 523-533.

Abdelmaksoud, I. R., Shalaby, A., Mahmoud, A., Elmogy, M., Aboelfetouh, A., Abou El-Ghar, M., El-Baz, A. (2021). Precise identification of prostate cancer from DWI using transfer learning. *Sensors*, 21(11), 3664.

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S. (2012, April). The'K'in K-fold Cross Validation. In ESANN (Vol. 102, pp. 441-446).

Aslan, M. (2022). Derin Öğrenme Tabanlı Otomatik Beyin Tümör Tespiti. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 34(1), 399-407.

Bjurlin, M. A., Carroll, P. R., Eggener, S., Fulgham, P. F., Margolis, D. J., Pinto, P. A., ... & Turkbey, B. (2020). Update of the standard operating procedure on the use of multiparametric magnetic resonance imaging for the diagnosis, staging and management of prostate cancer. *The Journal of urology*, 203(4), 706-712.

Carlsson, S., Assel, M., Sjoberg, D., Ulmert, D., Hugosson, J., Lilja, H., & Vickers, A. (2014). Influence of blood prostate specific antigen levels at age 60 on benefits and harms of prostate cancer screening: population based cohort study. Bmj, 348.

Chavda, M., & Degadwala, S. (2024). Prostate Cancer Gleason Score Classification Using Transfer Learning Models. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 10(2), 450-458.

Dorak, M. T., & Karpuzoglu, E. (2012). Gender differences in cancer susceptibility: an inadequately addressed issue. *Frontiers in Genetics*, 3, 268.

Fırıldak, K., & Talu, M. F. (2019). Evrişimsel sinir ağlarında kullanılan transfer öğrenme yaklaşımlarının incelenmesi. *Computer Science*, 4(2), 88-95.

Geert, L., Oscar, D., Jelle, B., Nico, K., and Henkjan, H.. ProstateX Challenge data, The Cancer Imaging Archive. (2017). https://www.kaggle.com/datasets/tgprostata/transverse-plane-prostate-dataset Accessed 02.05.2024.

Hamm, C. A., Baumgärtner, G. L., Biessmann, F., Beetz, N. L., Hartenstein, A., Savic, L. J., Penzkofer, T. (2023). Interactive explainable deep learning model informs prostate cancer diagnosis at MRI. *Radiology*, 307(4), e222276.

Himmerich, H., Kan, C., Au, K., & Treasure, J. (2021). Pharmacological treatment of eating disorders, comorbid mental health problems, malnutrition and physical health consequences. *Pharmacology & Therapeutics*, 217, 107667.

Hoar, D., Lee, P. Q., Guida, A., Patterson, S., Bowen, C. V., Merrimen, J., Clarke, S. E. (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.

Kanna, G. P., Kumar, S. J., Parthasarathi, P., & Kumar, Y. (2023). A review on prediction and prognosis of the prostate cancer and gleason grading of prostatic carcinoma using deep transfer learning based approaches. *Archives of Computational Methods in Engineering*, 30(5), 3113-3132.

Kılıçarslan, S., & Pacal, I. (2023). Domates Yapraklarında Hastalık Tespiti İçin Transfer Öğrenme Metotlarının Kullanılması. *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 5(2), 215-222.

Mahesh, T. R., Geman, O., Margala, M., & Guduri, M. (2023). The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification. *Healthcare Analytics*, 4, 100247.

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In Multivariate statistical machine learning methods for genomic prediction (pp. 109-139). Cham: Springer International Publishing.

Özbay, E., & Özbay, F. A. (2021). Derin Öğrenme ve Sınıflandırma Yaklaşımları ile BT görüntülerinden Covid-19 Tespiti. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 12(2), 211-219.

Özbay, F. A. (2023). A modified seahorse optimization algorithm based on chaotic maps for solving global optimization and engineering problems. *Engineering Science and Technology, an International Journal*, 41, 101408.

Seyyarer, E., Ayata, F., Uçkan, T., & Karci, A. (2020). Derin öğrenmede kullanilan optimizasyon algoritmalarinin uygulanmasi ve kiyaslanmasi. *Computer Science*, 5(2), 90-98.

Srivenkatesh, M. (2020). Prediction of prostate cancer using machine learning algorithms. *Int. J. Recent Technol. Eng*, 8(5), 5353-5362.

Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics*, 75, 34-46.

Tsuneki, M., Abe, M., & Kanavati, F. (2022). Transfer learning for adenocarcinoma classifications in the transurethral resection of prostate whole-slide images. *Cancers*, 14(19), 4744.

Weiss, K. R., & Khoshgoftaar, T. M. (2016, November). An investigation of transfer learning and traditional machine learning algorithms. In 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 283-290). IEEE.

Yuan, Y., Qin, W., Buyyounouski, M., Ibragimov, B., Hancock, S., Han, B., Xing, L. (2019). Prostate cancer classification with multiparametric MRI transfer learning model. *Medical Physics*, 46(2), 756-765.

Zainudin, Z., Shamsuddin, S. M., & Hasan, S. (2020). Deep layer CNN architecture for breast cancer histopathology image detection. In The International Conference on Advanced Machine Learning Technologies and Applications (Amlta2019) 4 (pp. 43-51). Springer International Publishing.

Zhong, X., Cao, R., Shakeri, S., Scalzo, F., Lee, Y., Enzmann, D. R., Sung, K. (2019). Deep transfer learning-based prostate cancer classification using 3 Tesla multi-parametric MRI. *Abdominal Radiology*, 44, 2030-2039.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43-76.