




RESEARCH ARTICLE

DETECTION OF CERVICAL CANCER FROM UTERINE CERVIX IMAGES USING
TRANSFER LEARNING ARCHITECTURES

Hanife GÖKER ^{1,*}

¹ Health Services Vocational College, Gazi University, 06830, Ankara, Türkiye
gokerhanife@gazi.edu.tr -  [0000-0003-0396-7885](https://orcid.org/0000-0003-0396-7885)

Abstract

Cervical cancer is a common and serious cancer affecting more than half a million women worldwide. For cervical cancer disease management, prognosis prediction, or optimizing medical intervention, early detection of the disease is critical. It is one of the types of cancer that can be successfully treated, as long as it is diagnosed early and managed effectively. In this study, an image processing-based solution was proposed for the diagnosis of cervical cancer from uterine cervix images using transfer learning architectures to reduce the workload and assist the experts. The proposed transfer learning model was tested using a publicly available dataset, which includes 917 uterine cervix images. Uterine cervix images were enhanced and brightness level using the histogram equalization method and denoised using the Gaussian filter. Then, the performances of AlexNet, DenseNet201, MobilenetV2, Resnet50, Xception, and VGG19 transfer learning architectures were compared. The transfer learning model performance was evaluated using the 10-fold cross-validation method. VGG19 transfer learning algorithm had the highest performance. VGG19 transfer learning algorithm achieved 98.26% accuracy, 0.9671 f1-measure, 0.9896 specificity, 0.9631 sensitivity, 0.9711 precision, 0.9552 Matthews correlation coefficient (MCC), and 0.955 kappa statistic. The combination of histogram equalization, Gaussian filter, and the VGG19 transfer learning approach can be used for accurate and efficient detection of cervical cancer from uterine cervix images. In this study, more accuracy was achieved compared to the known related studies in the literature.

Keywords

Image processing,
Transfer learning,
Deep learning,
Cervix cancer,
Gaussian filter

Time Scale of Article

Received: 20 June 2023
Accepted: 01 March 2024
Online date :28 June 2024

1. INTRODUCTION

Cervical cancer is a serious and common disease and affects the cervix, an important element of a woman's reproductive system. The disease is the fourth most typical cause of disease mortality among women worldwide. Every year, 604 000 women are thought to be diagnosed with cervical cancer, and about 342 000 of them pass away from the disease [1]. Developing countries have much higher mortality and morbidity rates compared to developed countries [2]. The health burden associated with this inequitable rate is expected to intensify, given that cervical cancer death rates are predicted to increase by about 22% between 2015 and 2030 [3]. Moreover, these cancer cells can spread to other organs including the liver, rectum, lungs, and bladder [4]. Preventable if caught early, cervical cancer is becoming an increasingly common disease of inequality and lack of access to healthcare due to shortcomings in accountable social and health systems [5].

In clinical practice, detection of tumor biomarkers, pap-smear tests, colposcopy, and medical imaging techniques are among the examinations for the diagnosis of cervical cancer [6]. Cervical cancer has a

*Corresponding Author: gokerhanife@gazi.edu.tr

high cure rate if found at an early stage [7]. However, since cervical cancer is asymptomatic in its early stages, it is often undetectable and can only be identified by routine checkups or pelvic exams [8]. Therefore, regular scanning is imperative to reduce mortality rates. By detecting color changes in the cervix, the healthcare professional can identify the presence of potential malignant or pre-cancerous lesions [9]. Pre-cancer scanning has played a crucial role in reducing the mortality rate and incidence of cervical cancer in the last fifty years. Unfortunately, because of the increased workload, vision scanning results in misdiagnosis and ineffective diagnostic methods [10]. In addition, significant regional and global disparities in cervical cancer outcomes have prompted international gynecological cancer societies to pursue evidence-based methods aimed at improving the quality of care for patients [11]. Cervical cancer is frequently detected using three-dimensional (3D) medical imaging techniques including computed tomography (CT), nuclear magnetic resonance imaging (MRI), and positron emission tomography (PET). 3D images may include invaluable information on medical results [12]. The pap-smear test detects changes in cervical cells and shows whether the cells are prone to cancer. It requires a small scraping of cervical tissue and analysis in a lab. Sample cell images are collected in a container with additional liquid and these cell images are manually analyzed by a qualified pathologist. This procedure is time-consuming and tedious given the total number of specialist pathologists [13]. However, specialist pathologists with the ability to diagnose images often require long-term training, and the diagnosis of images is a subjective, uncertain process, and its accuracy is greatly influenced by the experience and condition of the experts [6]. In addition, it takes a lot of time to analyse hundreds of images for 3D medical imaging analysis, and the intensive image scanning job has greatly hampered clinicians [14]. For the diagnosis of cervical cancer in practice; computer-aided diagnosis systems that can automatically analyze medical images are needed due to reasons such as the need for routine checks, pelvic examinations, and expert pathologists [9], the high workload of specialists [10], and the diagnostic procedure being time-consuming, boring, and subjective [13]. Transfer learning-based models have a great capacity to automatically process complex and very large image data and the models are strong candidates to overcome these limits. Transfer learning models store the information obtained in solving any problem and use that information when faced with another problem. Features and weights obtained from previously trained models are used for new tasks with transfer. Thus, by using previous knowledge with transfer, models with higher success and faster learning are obtained with less training data.

In recent years, machine learning and deep learning methods have gained popularity for the detection of cervical cancer. Ahishakiye et al. (2020) performed a classification model for the detection of cervical cancer using the ensemble learning approach. The model achieved 87.21% accuracy [15]. Chandran et al. (2021) proposed an ensemble deep learning-based classification model to detect cervical cancers from colposcopy images. They adopted VGG19 and colposcopy ensemble network (CYENET) algorithms. VGG19 algorithm achieved 73.30% accuracy and CYENET algorithm achieved 92.30% accuracy [10]. Saini et al. (2020) presented a deep-learning-based method for detection of cervix cancer from colposcopy images. They compared transfer learning models, including LeNet, VGG16, DenseNet, AlexNet, ResNet50, GoogleNet, and ColpoNet, for classifying cervical cancer. The ColpoNet algorithm classified cervical cancer with 81.353% accuracy [16]. Similarly, Priyanka and Raju (2021) developed deep learning models to automatically classify cervical cancer. They performed image preprocessing such as resizing the images, converting the images into grayscale images, and expanding the dimensions of the images. Then, their proposed study classified cervical cancer and predicted it with an accuracy of 74.04% using ResNet50 algorithm [17]. Mehmood et al. (2021) proposed a hybrid approach that combines Random Forest (RF) and shallow neural networks for cervical cancer detection. The hybrid approach used the Pearson correlation between the input variables and the output variable to preprocess the data and achieved an accuracy of 93.6% [18]. Alsmariy et al (2020) used the ensemble voting classifier by combining Decision Tree (DT), Logistic Regression (LR), and RF machine learning algorithms to find a model that can diagnose cervical cancer with high accuracy and sensitivity. They used a cervical cancer dataset from the University of California at Irvine (UCI). The highest accuracy (97.44%) was achieved using SMOTE, voting, and Principal Component Analysis (PCA) [19]. Kalbhor et al (2023) used the publicly available Herlev pap-smear dataset. In the study,

features were extracted using pre-trained convolution neural networks such as Alexnet, Googlenet, Resnet18, and Resnet50. The cervix images were classified with machine learning algorithms, namely simple logistic, RBFNetwork, BayesNet, Naive Bayes (NB), RF, random tree, and decision table. The highest accuracy of 96.01% was achieved using GoogleNet [20]. Early detection of cervical cancer is critical in the treatment of the disease, prognosis prediction, and optimizing the treatment. If it is identified early and treated appropriately, cervical cancer is among the most successfully treatable forms of cancer. However, there are deficiencies in its correct diagnosis. Although there is increasing interest in deep learning-based models for automated analysis of image data, further studies are required to make deep learning-based models reliable and practical for clinical use. Studies investigating early diagnosis of cervical cancer from medical images are limited. There is a need for new studies to be conducted, to compare the results, and to improve accuracy.

In the study, we proposed an image processing-based solution to detect cervical cancer from uterine cervix images using transfer learning architectures to assist the experts and reduce the workload of the experts. The significant contributions of this study are summarized below.

- a) An image processing-based solution model was proposed for the detection of cervical cancer from uterine cervix images.
- b) The best classification performance model was chosen after comparing the transfer learning architectures, namely AlexNet, DenseNet201, MobilenetV2, Resnet50, Xception, and VGG19.
- c) In this study, more accuracy was achieved compared to the known related studies in the literature.
- d) Early diagnosis of cervix cancer from uterine cervix images offers opportunities for early intervention and effective management of the process.

2. MATERIAL AND METHODS

2.1. Proposed Model

The proposed model contains the following implementation phases: i) Firstly, the image preprocessing was performed on the uterine cervix images. In the image preprocessing phase, the images were resized, the images were sharpened using histogram equalization, and the noise on the images was removed using the Gaussian filter, ii) Then, the classification successes of the transfer learning architectures were compared. iii) The 10-fold cross-validation was adopted to assess the proposed model performance, due to the low estimation bias and error of this method. iv) Finally, the model evaluation metrics of the transfer learning architectures were calculated, and the transfer learning architecture with the highest performance was adopted. The implementation phases of the proposed model are given in Figure 1.

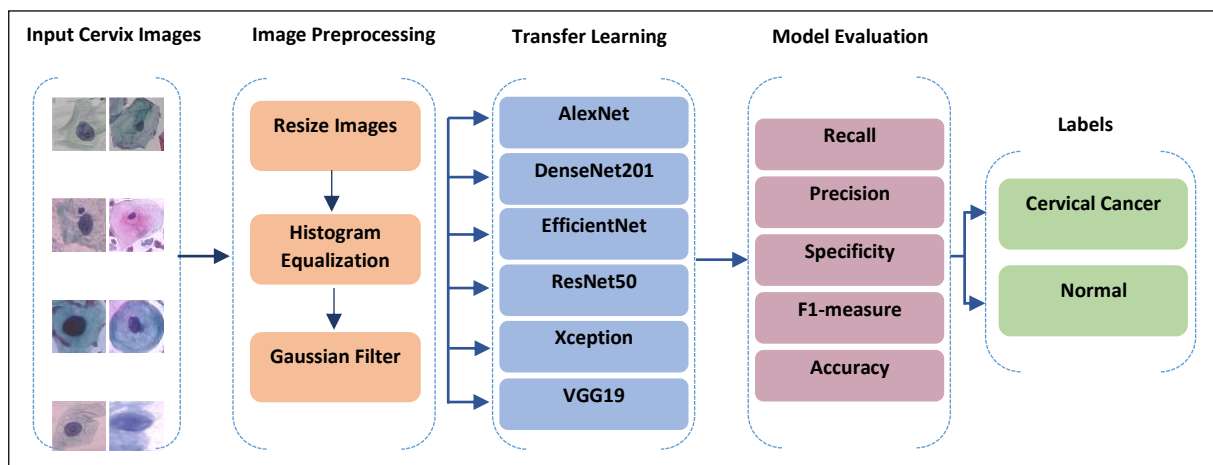


Figure 1. The implementation phases of the proposed transfer learning model

2.2. Dataset

The study was conducted on the publicly available Herlev pap-smear dataset [21]. The image dataset was gathered by the Technical University of Denmark and Herlev University Hospital. It consisted of 917 uterine cervix images. 242 images were normal cells and 675 images were cervix cancer cells. The pap-smear image has a resolution of 0.201 f.J.m per pixel. As part of the smear inspection, sample uterine cervix tissues were used to create the dataset [21]. The dataset has 7 cell types. The distribution of the image dataset is given in Table 1:

Table 1. The distribution of the dataset

Number	Cell Type	Counts	Category
1	Superficial epithelial	74	Normal cells (242 images)
2	Intermediate epithelial	70	
3	Columnar epithelial	98	
4	Mild squamous non-keratinizing dysplasia	182	Cervix cancer cells (675 images)
5	Moderate squamous non-keratinizing dysplasia	146	
6	Severe squamous non-keratinizing dysplasia 1	197	
7	Squamous cell carcinoma in sti intermediate	150	

In this study, the dataset was divided into 2 categories (normal and cervical cancer) and focused on the detection of cervical cancer from the uterine cervix images. Samples of cervix cancer images and normal cervix images are given in Figure 2 and Figure 3:

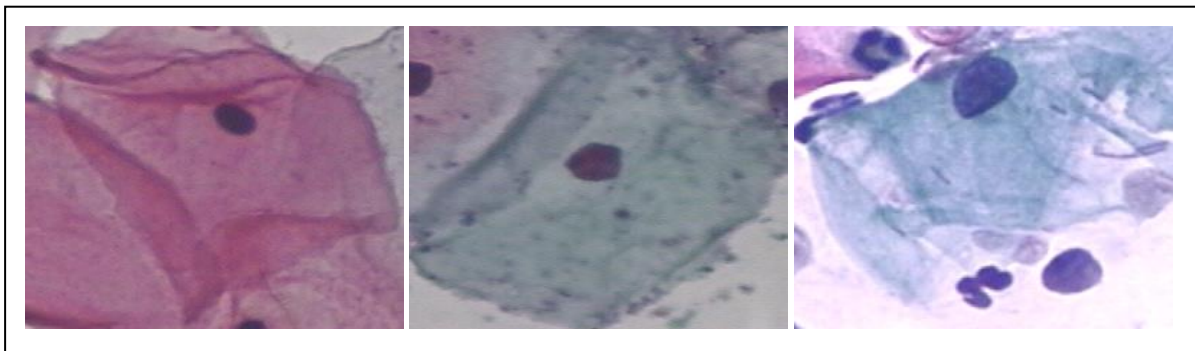


Figure 2. Samples of normal cervix image

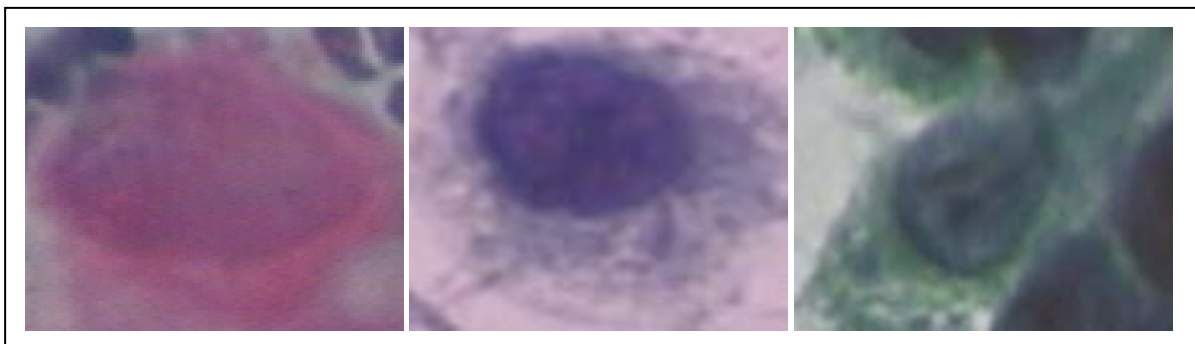


Figure 3. Samples of cervix cancer image

2.3. Image Preprocessing

In the image preprocessing stage, the images were converted to .jpeg format and resized according to the default input sizes used by the transfer learning architectures. Uterine cervix images resized to 227x227x3 for AlexNet, 224x224x3 for DenseNet201, 224x224x3 for MobilenetV2, 224x224x3 for RestNet50, 224x224x3 for Xception, and 224x224x3 for VGG19. Then, at this stage, histogram equalization was used for image enhancement, and a Gaussian filter was used to reduce noise on the images. After applying histogram equalization, the samples of cervix images are given in Figure 4:

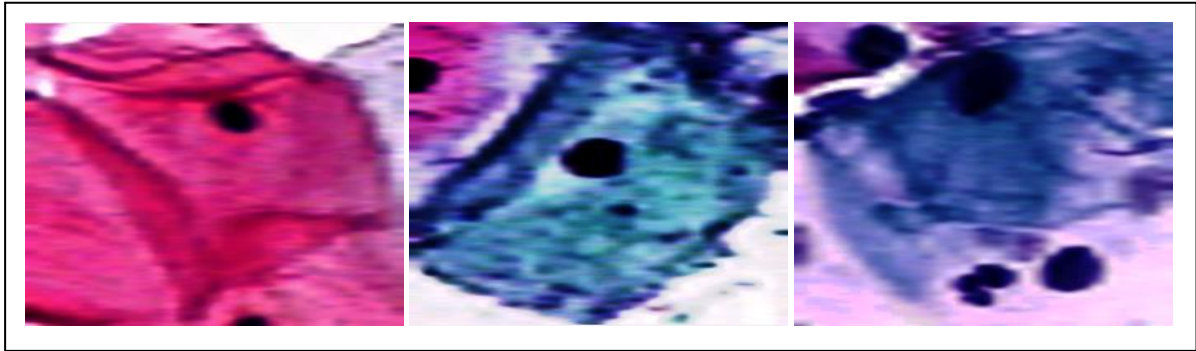


Figure 4. Samples of cervix images after applying histogram equalization

Image enhancement might be considered as one of the critical steps for image analysis. The image enhancement aims to improve an image's quality to make it better suited for a certain application. Histogram equalization is a popular and straightforward image enhancement method. It aims to have the output image histogram in a uniform distribution. By ensuring that there are approximately the same number of pixels for each brightness level, the brightness level of the image is adjusted, thereby sharpening the image [22]. It deals with the overall appearance of an image and shows the relationship between the quantity of pixels and each intensity value. It provides the likelihood of occurrence of a pixel with a certain density. In the histogram equalization method, during the whole dynamic density range, the pixels are evenly dispersed. The general formula of the histogram equalization is given below [23]:

$$p_r(r_k) = \frac{n_k}{n} \quad k=0, 1, 2, \dots, L-1 \quad (1)$$

$p(r_k)$ shows the ratio of the k.th tone value represented in the image, n_k shows the number of k.th tone in the image, and n shows the total number of pixels. In histogram equalization, firstly the s_k cumulative probability is calculated as in equation 2. Then, by performing an inverse transformation as in equation 3, what color tone will replace is calculated.

$$s_k = T(r_k) = \sum_{j=0}^k p_r(r_j) = \sum_{j=0}^k \frac{n_j}{n} \quad k=0, 1, 2, \dots, L-1 \quad (2)$$

$$\begin{aligned} r_k &= T^{-1}(s_k), & 0 \ll s_k \ll 1 \\ T^{-1}(s_k) &= (L - 1) * T(r_k), & k=0, 1, 2, \dots, L-1 \end{aligned} \quad (3)$$

The samples of cervix images after applying the gaussian filter are given in Figure 5:

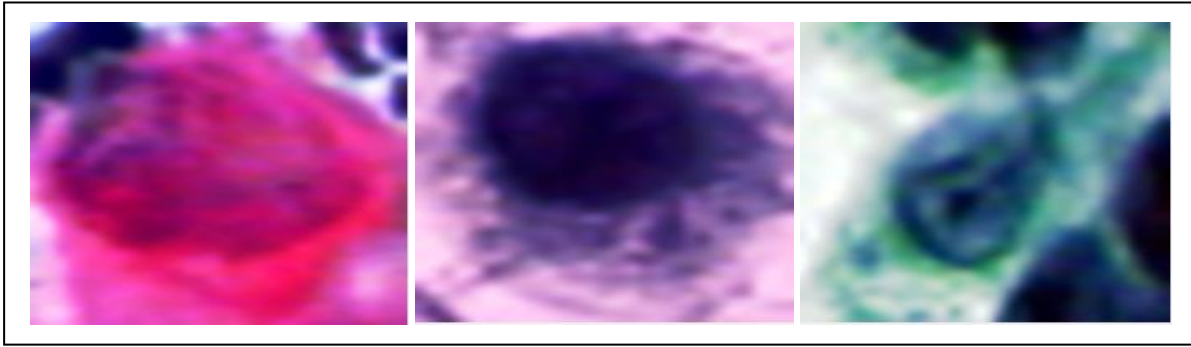


Figure 5. Samples of cervix images after applying the gaussian filter

The reason for using the Gaussian filter in this study is to protect the spatial-spatial frequency components on the images, that is, it protects the changes on the images at very frequent intervals. Gauss subtracts the weighted average of each pixel region. The weighting increases as you get closer to the center pixel value. This filter provides a finer trim, unlike other filters, and preserves edges better than other filters of a similar size. Most other image filters behave like a low-pass filter, which adopts the all-or-nothing rule and exhibits oscillations as a frequency response. The Gaussian filter, on the other hand, does not oscillate. The general formula of the Gaussian filter is given below [24]:

$$G(r) = \frac{1}{(2\pi\sigma^2)^{N/2}} e^{-r^2/(2\sigma^2)} \quad (4)$$

2.4. 10-Fold Cross-Validation

The k-fold cross-validation is used to assess whether the model performance is random and has an overfitting problem. The dataset is split into k separate subgroups. The k value is usually chosen as 5 or 10. In this study, the k value was adopted as 10. One of the groups serves as the test set, and the other groups serve as the training set. The model is trained and tested with the other group by iteratively repeating each group. The overall performance and error rate of the model are calculated as the average of the 10 results obtained. This minimizes deviations and mistakes from dispersion and fragmentation and reduces prediction bias by allowing each part to be utilized alternately for both training and testing. However, training and testing the model for each k requires overhead and time. The 10-fold cross-validation was used in this study because it overcomes overfitting, minimizes prediction bias and error, and enables the model to be trained using several training-test groups. The performances of AlexNet, DenseNet201, MobilenetV2, ResNet50, Xception, and VGG19 transfer learning architectures were compared.

2.5. Transfer Learning

Transfer learning architectures are a type of machine learning where a model created for another task is used to solve different or similar problems. It is frequently simpler and faster to use a pre-trained network with transfer learning than training a network from scratch. The most important ones are network speed, accuracy, and size. The tradeoff between these characteristics should be taken into account when choosing a network. These architectures can be used for different purposes such as classification or feature extraction. In this study, AlexNet, DenseNet201, MobilenetV2, Resnet50, Xception, and VGG19 transfer learning architectures were used for classification tasks. Different experiments were performed on the selection of hyper-parameters for transfer learning architectures whose performances were compared in the study, and optimum hyper-parameters were discovered. In the experiments, “sgdm”, “adam”, and “rmsprop” were performed for the “optimizer” hyper-parameter in transfer learning architectures and the “sgdm” value, which achieved the best performance, was selected. The

“MiniBatchSize” hyper-parameter was tried as “16” and “32” values and the “16” value, which achieved the best performance, was selected. Also, “30” and “60” were performed for the “MaxEpochs” hyper-parameter, and the “30” value, which achieved the best performance, was selected. The transfer learning architectures and optimizer parameters used are given in Table 2.

Table 2. Transfer learning architectures and optimizer parameters

Hyper-Parameters	AlexNet	DenseNet201	MobilenetV2	ResNet50	Xception	VGG19
Input size	227x227x3	224x224x3	224x224x3	224x224x3	224x224x3	224x224x3
Optimizer	sgdm	sgdm	sgdm	sgdm	sgdm	sgdm
MiniBatchSize	16	16	16	16	16	16
MaxEpochs	30	30	30	30	30	30
InitialLearnRate	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4
ValidationFrequency	3	3	3	3	3	3
Momentum	0.9	0.9	0.9	0.9	0.9	0.9
LearnRateDropPeriod	10	10	10	10	10	10
L2Regularization	0.005	0.005	0.005	0.005	0.005	0.005
WeightLearnRateFactor	20	20	20	20	20	20
BiasLearnRateFactor	20	20	20	20	20	20

2.5.1. AlexNet

AlexNet is a fundamental, effective, and simple transfer learning architecture. It mainly consists of five convolutional layers; the first layer to the fourth layer, followed by the pooling layer, and the fifth layer, followed by three fully connected layers. The image input size is $227 \times 227 \times 3$ in AlexNet transfer learning architecture. The structure of the AlexNet is given in Figure 6 [25].

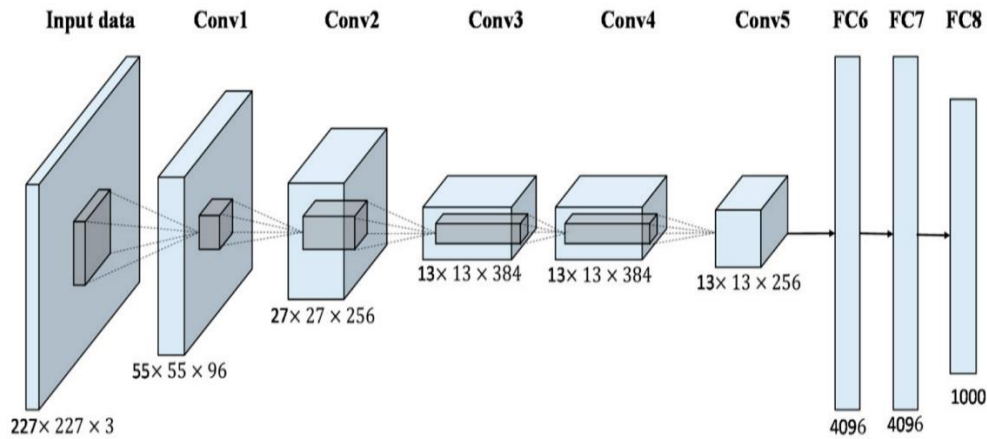


Figure 6. The AlexNet architecture.

2.5.2. DenseNet201

The DenseNet201 architecture achieves the highest possible amount of information flow density between network layers. Because the feature maps produced by the input from the previous layers are transmitted to the next layer by each layer. The biggest advantage of this architecture is that it guarantees the reuse of the features acquired by making sure that the features created in each layer pass on to the following layers. In this architecture, all layers are related to one another. Its name is a dense connected convolutional network as a result. A DenseNet architecture with L layers has $L(L+1)/2$ direct connections. For each layer, the feature maps of all previous layers are used as input. The image input

size is 224 x 224 x 3 in DenseNet201 transfer learning architecture. The structure of the DenseNet is given in Figure 7 [26].

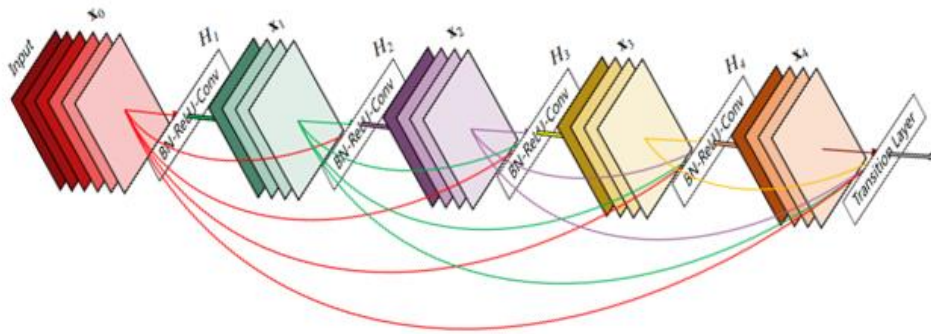


Figure 7. The DenseNet architecture (5-layer dense block)

2.5.3. MobilenetV2

MobileNet is a simple deep convolutional neural network model that is much smaller in size and faster in performance than many other popular models. In depth separable convolutions, a single filter is applied to each input channel. A batchnorm and rectified linear unit (ReLU) nonlinear activation layer comes after all layers, except the fully connected layer that feeds the softmax layer to classify the output. MobileNet has 28 layers except for depth and point convolutions [27]. The image input size is 224 x 224 x 3 in the MobilenetV2 transfer learning architecture. The structure of the MobilenetV2 is given in Figure 8 [28].

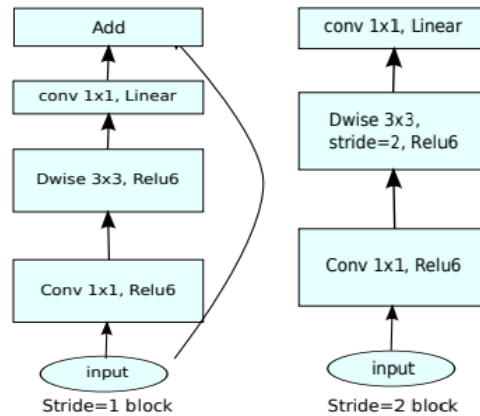


Figure 8. The MobilenetV2 architecture

2.5.4. ResNet50

In the ResNet50 architecture, allowing the transition to the lower layers by ignoring the changes in some layers has increased the success rates. There is information about a 50-layer network structure and the connections between layers in the Resnet50 architecture. ResNet architecture uses convolution layers (1x1), (3x3), (1x1) instead of using 2 (3x3) convolutions. The structure of the ResNet50 is given in Figure 9 [29]. The image input size is 224 x 224 x 3 in ResNet50 transfer learning architecture.

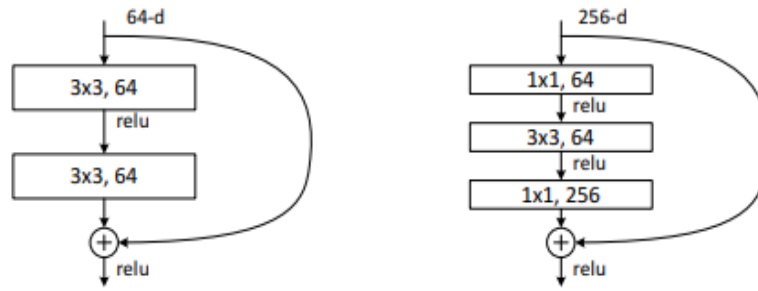


Figure 9. The ResNet architecture

2.5.5. Xception

The Xception architecture has been a developing network by building on top of the InceptionV3 network. In this architecture, the convolutional part is different from other architectures. A classic network moves a filter over multidimensional matrices such as height, width, and depth in the convolutional part. In this convolutional part, Xception network provides two different approaches in addition to the developments in InceptionV3. These are pointwise convolution and depthwise convolution. In the depthwise convolution part, it reaches the result by processing only one channel, not every channel. It is also called “Extreme Inception”. The addition it adds to InceptionV3 is that 1x1 convolution uses convolution for each of its output channels. Figure 10 shows Inception architecture and Extreme Inception architecture [30]. The image input size is 299 x 299 x 3 in Xception transfer learning architecture.

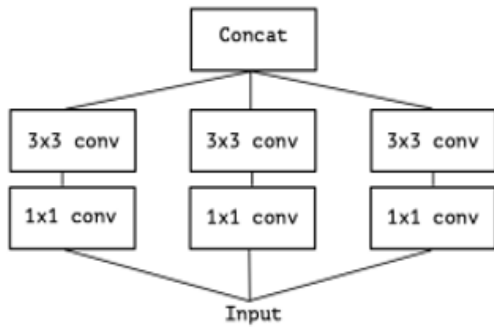


Figure 10.a. Simplified Inception module

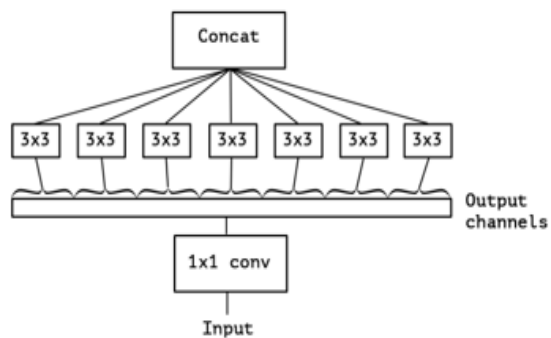


Figure 10.b. Extreme Inception module

The Xception architecture includes depthwise separable convolutions. As an alternative to classical convolutions, which are assumed to be much more efficient in terms of computation time, depthwise separable convolutions are offered. It is not only concerned with spatial dimensions, but also with the dimension of depth and the number of channels. Depthwise separable convolutions split a core (filter) into two different cores that do two separate convolutions. While one of these kernels is deeply convoluted, the other is point convolution.

2.5.6. VGG19

VGG is a classical convolutional neural network architecture with multiple layers and was developed to enhance the performance of the model by increasing the depth of convolutional neural networks. “16” and “19” represent the number of convolutional layers in the model. VGG19 architecture is a variant of the VGG model which in short consists of 19 layers. This means that VGG19 has three more convolutional layers than VGG16. The architecture includes 16 convolutions, 3 fully connected layers,

5 maxpools layers, and 1 softmax layer. The structure of the VGG19 is given in Figure 11 [31]. The image input size is 224 x 224 x 3 in VGG19 transfer learning architecture.

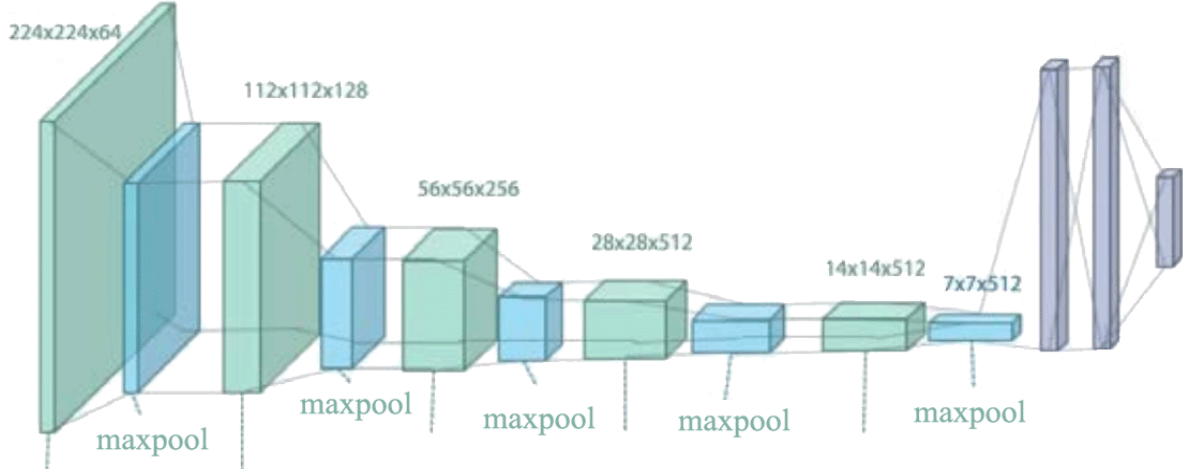


Figure 11. The VGG19 architecture

2.6. Performance evaluation metrics

Performance evaluation metrics, namely accuracy, f1-measure, specificity, sensitivity, precision, MCC, and kappa statistics are widely used to evaluate the performance of classification models [32]. These performance evaluation metrics are calculated using the confusion matrix, which is a table containing four different parameters of predicted and actual samples. These are false positive (FP), true positive (TP), false negative (FN), and true negative (TN). The mathematical formulas for these metrics are given below:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (5)$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (6)$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (7)$$

$$\text{MCC} = (\text{TN} \times \text{TP} - \text{FP} \times \text{FN}) / \sqrt{(\text{FN} + \text{TN}) \times (\text{FP} + \text{TP}) \times (\text{TN} + \text{FP}) \times (\text{TP} + \text{FN})} \quad (8)$$

$$\text{F1-measure} = (2 \times \text{Precision} \times \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity}) \quad (9)$$

$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN}) \quad (10)$$

Kappa statistic is calculated by following the mathematical formula steps below [33]:

$$P_o = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

$$P_1 = [(\text{FN} + \text{TP}) / (\text{FP} + \text{FN} + \text{TP} + \text{TN})] \times [(\text{FP} + \text{TP}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN})]$$

$$P_2 = [(\text{TN} + \text{FP}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})] \times [(\text{FN} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})]$$

$$P_e = P_1 + P_2$$

$$\text{Kappa statistic} = (P_o - P_e) / (1 - P_e) \quad (11)$$

3. EXPERIMENTAL RESULTS AND DISCUSSION

In the study, a deep learning model for the diagnosis of cervical cancer was proposed using uterine cervix images. The dataset contains 917 images, of which 675 are cancer images and 242 are normal. The images were resized according to the input image size appropriate for each transfer learning architecture. The images were sharpened and brightness levels were adjusted using histogram equalization. Gaussian filters were used to reduce noise in images. The performances of AlexNet, DenseNet201, MobilenetV2, Resnet50, Xception, and VGG19 transfer learning architectures were compared. The model performance was evaluated using the 10-fold cross-validation. The total dataset was divided into 10 folds in this method. For each fold, 9 parts were used alternately for training and 1 part for testing. Thus, all images were evaluated as both training and test datasets. While evaluating the performance of the model, the average of 10 folds was calculated. 10-fold cross-validation is generally the preferred method because it allows training the proposed model with more than one training-test group, thus reducing the prediction bias. The performances of the transfer learning architectures were compared. The confusion matrix of the architectures, showing the number of images included in the correct label and the number of images included in the incorrect label, are given in Table 3.

Table 3. The confusion matrix of the transfer learning architectures

Transfer Learning Architectures	Labels	Confusion matrix parameters		Classified	
		Normal	Cervical Cancer	Correctly (TP+TN)	Incorrectly (FP+FN)
AlexNet	Normal	216	26	873	44
	Cervical Cancer	18	657		
DenseNet201	Normal	224	18	872	45
	Cervical Cancer	27	648		
MobilenetV2	Normal	219	23	870	47
	Cervical Cancer	24	651		
ResNet50	Normal	215	27	866	51
	Cervical Cancer	24	651		
Xception	Normal	206	36	855	62
	Cervical Cancer	26	649		
VGG19	Normal	235	7	901	16
	Cervical Cancer	9	666		

When Table 3 was examined, the VGG19 transfer learning architecture had the highest number of correctly classified images. VGG19 transfer learning architecture was followed by AlexNet, DenseNet201, MobilenetV2, ResNet50, and Xception, respectively. In the confusion matrix of VGG19 transfer learning architecture, the total number of incorrectly classified images (FN+FP) was 16 and the total number of correctly classified (TN+TP) images was 901. The model evaluation metrics, which are accuracy, f1-measure, specificity, sensitivity, precision, MCC, and kappa statistics are calculated using the parameters in the confusion matrix, which are FP, TP, FN, and TN. The success of the models is evaluated using these model evaluation metrics. The performance analysis results of the transfer learning are given in Table 4:

Table 4. The performance analysis results of the transfer learning architectures

Architectures	Model Evaluation Metrics						
	Sensitivity	Precision	Specificity	MCC	F1-measure	Kappa	Accuracy
AlexNet	0.9231	0.8926	0.9619	0.8754	0.9076	0.875	0.9520
DenseNet201	0.8924	0.9256	0.9730	0.8755	0.9087	0.875	0.9509
MobilenetV2	0.9012	0.9050	0.9659	0.8683	0.9031	0.868	0.9487
ResNet50	0.8996	0.8884	0.9602	0.8563	0.8940	0.856	0.9444
Xception	0.8879	0.8512	0.9474	0.8240	0.8692	0.824	0.9324
VGG19	0.9631	0.9711	0.9896	0.9552	0.9671	0.955	0.9826

When the performance analysis results of the architectures are examined in Table 4, the VGG19 transfer learning architecture has the highest success (accuracy = 98.26%). The performance analysis results of VGG19 architecture were calculated as 0.9671 f1-measure, 0.9896 specificity, 0.9631 sensitivity, 0.9711 precision, 0.9552 MCC, and 0.955 kappa statistic. Model performance criteria should be close to 1. If performance evaluation metrics are close to 1, it proves that the model does not have an accidental success. In addition, the kappa statistic between 0.81-1.00 shows an almost “perfect match”. Kappa statistic values of all transfer learning architectures were above 0.81. While evaluating the success of the model, the f1-measure should be examined together with the accuracy. The accuracy is the correct predictions divided by the total number of datasets. The f1-measure is the harmonic mean of the sensitivity and precision. Accuracy may not produce reliable results if the intergroup distribution is not balanced in the data set, but the f1-measure produces accurate results even if the intergroup distribution is not even. For this reason, it is more appropriate to present the f1-measure together with the accuracy in the analysis of the models. The f1-measure values of transfer learning architectures were close to 1. In Figure 12, the accuracy and f1-measure of transfer learning architectures are given.

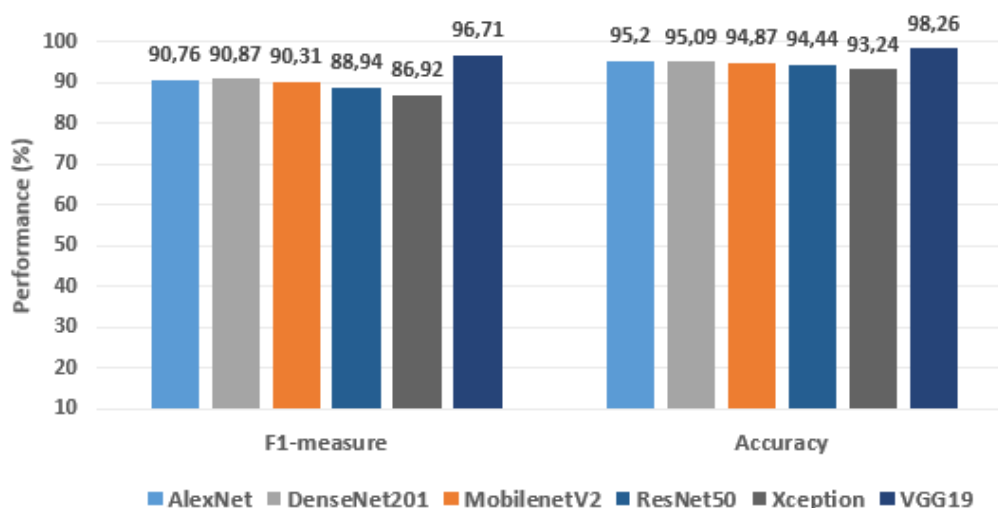


Figure 12. The accuracy and f1-measure of transfer learning architectures

The performance results of the VGG19 transfer learning architecture, which had the highest performance, were compared with the related studies that included cervical cancer detection from cervical images. The comparative analysis is given in Table 5.

Table 5. Comparative analysis of related studies using the same dataset about cervical cancer detection

Researchers	Data Preprocessing	Classifier	Accuracy
Arora et al. (2021) [4]	Geometrical features, Texture features, Feature selection with PCA	Polynomial SVM	95%
		Gaussian SVM	85%
		Quadratic SVM	85%
Priyanka & Raju (2021) [17]	Resize, convert to grayscale, expand the dimensions of the images	ResNet50	74.04%
Sun et al. (2017) [34]	11 nucleus features and 9 cytoplasm features. Feature selection with Relief	NB	92.04%
		C4.5	92.80%
		LR	93.13%
		RF	94.44%
Malli & Nandyal (2017) [35]	Colour quantization, fuzzy c, detected nucleus, nucleus marked area	kNN	88.04%
		ANN	54%
Nguyen et al (2019) [36]	Ensemble of Resnet152, Inception-Resnet-v2 and Inception-v3, feature concatenation	InceptionV3+ Resnet152+ Inception-Resnet-v2	93.04%
Lin et al (2019) [37]	Data augmentation, cell morphology and image patch extraction	GoogLeNet-5C	94.5%
			71.3%
			64.5%
Khamparia et al. (2020) [38]	Feature extraction with convolutional neural network encoder (ResNET50)	kNN	97.67%
		NB	96.08%
		LR	93.45%
		RF	97.89%
		SVM	97.44%
Ravindran et al. (2021) [39]	Reduce sounds with median sensor, Boost comparison with CLAHE	Ensemble	95.12%
Lavanya Devi & Thirumurugan (2022) [40]	Feature extraction with modified fuzzy c-means, Data augmentation with PCA	KNN	94.86%
		SVM	79.89%
		RF	86.41%
Shinde et al. (2022) [41]	PCA and machine learning ensemble	FCNN	87.50%
		Voting Ensemble	87.90%
		ANN	97.95%
		AlexNet	95.20%
		DenseNet201	95.09%
		MobilenetV2	94.87%
The Proposed Method	Resize images, Histogram equalization, Gaussian Filter	ResNet50	94.44%
		Xception	93.24%
		VGG19	98.26%

The performance analysis results of this study were discussed with the related studies according to data preprocessing, feature extraction and feature selection methods, classification algorithms, and accuracy. The histogram equalization method was used to enhance and sharpen the images, and the Gaussian filter to reduce noise in this study. When the methods used in related studies in the literature are examined in Table 5, geometrical features and texture features [4], nucleus features and cytoplasm features [34], image patch and cell morphology extraction [37], convolutional neural network encoder [38] and modified fuzzy c-means [40] were used for feature extraction. PCA [4] and Relief [34] methods were used for feature selection. Color quantization [35], resize images, and convert to grayscale [17] were used for data pre-processing. The median sensor was used to reduce sounds, and CLAHE was used to boost comparison [39]. In this study, Gaussian filter was adopted to reduce noise. Because this filter, unlike other filters, does not adopt the all-or-nothing rule, it does not oscillate, preserves the edges better, and provides a finer trim [24]. Histogram equalization is one of the popular, fastest, and most straightforward image processing methods to improve contrast in images [23]. Therefore, histogram equalization was adopted to boost contrast in images in the study. In this study, the performances of AlexNet, DenseNet201, MobileNetV2, ResNet50, Xception, and VGG19 transfer learning architectures were analyzed. When the studies on the detection of cervical cancer are examined in Table 5, machine learning algorithms, namely NB [34, 38], LR [34, 38], RF [34, 38, 41], SVM [4, 38, 41], kNN [35, 38,

40], transfer learning algorithms, namely ResNet [17, 36], Inception [36], and ensemble learning algorithm [39, 41] are used. The accuracy of the VGG19 transfer learning architecture (98.26%) was higher than the other studies in Table 5 [34, 36, 38-41]. The VGG transfer learning architecture is a famous and widely used CNN that demonstrated for the first time that accurate image processing is possible with a deep network and small convolutional filters. VGG19 performs well in image processing and generalizes well to related tasks because a) it is trained on millions of images, b) it uses very small receptive fields (3x3 with a stride of 1 step), and c) the decision function is more distinctive because it has 3 ReLU units instead of just one. VGG transfer learning architecture pioneered in making the network deeper and the filter compact (3x3). A stack of three convolutional layers with three-layer filters has the same receptive space as one convolutional layer with seven layers but only requires 45% fewer parameters to train. Additionally, the activation is sharper than the single-layer design since the stack has three activation functions. As the network gets deeper, the same architecture (layer groups followed by pooling) is repeated. Deeper levels generally have more filters than those nearer the input. While the deeper layers might need to recognize the full image, the first layers only need to distinguish shapes like lines and edges [42-43]. Deep layers contain the largest number of convolutional filters, as there are more complex object combinations. A greater number of filters allows the network to recognize a variety of complex shapes and objects, and as a result, more filters enhance the performance of the VGG transfer learning architecture.

4. CONCLUSION

In conclusion, we compared the performance of transfer learning architectures for the detection of cervical cancer from uterine cervix images in this study. Firstly, the images were resized to the appropriate input size for all architectures. Then, the histogram equalization method was used to enhance the images, and the Gaussian filter was applied to reduce the noise on the images. Finally, the performances of AlexNet, DenseNet201, MobilenetV2, Resnet50, Xception, and VGG19 transfer learning architectures were compared. The models were evaluated according to model evaluation metrics using the 10-fold cross-validation method. The strengths of the proposed model are to boost comparison in images using the histogram equalization method, denoised from images using the Gauss filter method, reduce prediction bias and error using the 10-fold cross-validation method, and provide a high and satisfactory performance by comparing transfer learning architectures. Performance analysis results showed that the VGG19 transfer learning algorithm obtained higher success than other transfer learning algorithms. The VGG19 algorithm had 98.26% accuracy, 0.9671 f1-measure, 0.9896 specificity, 0.9631 sensitivity, 0.9711 precision, 0.955 kappa statistic, and 0.9552 MCC. The fact that performance evaluation metrics are very close to 1 proves that the model did not have an accidental success. In addition to the high accuracy in the test results, a kappa statistical value between 0.8-1.0 indicates a “perfect match”. The accuracy achieved with the VGG19 transfer learning algorithm was higher compared to known related studies.

There are several limitations to the study. Firstly, transfer deep learning algorithms require a large number of images. Obtaining images for biomedical image processing is a very difficult and time-consuming process. Therefore, a public database was used in this study and the results of the study are limited to this database. For future studies, it is recommended to combine and analyze different cervical cancer databases. Secondly, transfer learning algorithms need high-quality hardware. Therefore, it is very time-consuming to run many experiments repeatedly. For future studies, a CNN algorithm with fewer layers and a simpler structure can be designed instead of highly computational and multi-layer transfer learning architectures such as AlexNet, DenseNet201, MobilenetV2, Resnet50, Xception, or VGG19. The results of the designed CNN algorithm can be compared with the transfer learning algorithms. Despite these limitations, the proposed model has many strengths. A high-performance deep learning solution based on transfer learning is presented to support expert opinion for the detection of cervical cancer from uterine cervix images. The performance of the proposed model is high because the images have gone through the data preprocessing process, which includes using the histogram

equalization method to improve the images and implementing the Gaussian filter to reduce the noise in the images. Moreover, prediction bias and error were minimized and overfitting was overcome using 10-fold cross-validation in the study. Consequently, the practical contributions of the proposed model can be summarized as follows: (i) The proposed transfer learning-based model provides a solution to support expert decision and classify cervical cancer with high performance and low cost. (ii) This study constitutes a successful example of how transfer learning models can be used for the diagnosis of diseases. (iii) Besides, the study contributed practically to the development of diagnostic decision support systems. (iv) Using the proposed model in clinical studies minimizes subjectivity, reduces the workload of experts, and enables rapid decision-making. The transfer learning-based model can achieve a high classification performance for the detection of different diseases from images.

CONFLICT OF INTEREST

The author stated that there are no conflicts of interest.

CRedit AUTHOR STATEMENT

Hanife Göker: Conceptualization, Methodology, Software, Formal analysis, Validation, Visualization, Writing.

REFERENCES

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 2021; 71: 3, 209-249.
- [2] Siegel RL, Miller KD, Jemal A. *Cancer statistics 2019*. *CA: A Cancer Journal for Clinicians* 2019; 69: 1, 7-34.
- [3] World Health Organization. Projections of mortality and causes of death 2015 and 2030. Online available: https://www.who.int/health-topics/cervical-cancer#tab=tab_1. [Accessed: 02-January-2023]
- [4] Arora A, Tripathi A, Bhan A. Classification of cervical cancer detection using machine learning algorithms. In *2021 6th International Conference on Inventive Computation Technologies (ICICT) 2021*; pp. 827-835.
- [5] Freeman HP, Wingrove BK. *Excess cervical cancer mortality: a marker for low access to health care in poor communities*. National Cancer Institute, Center to Reduce Cancer Health Disparities Rockville, MD. 2005.
- [6] Li C, Xue D, Zhou X, Zhang J, Zhang H, Yao Y, Kong F, Zhang L, Sun, H. Transfer learning based classification of cervical cancer immunohistochemistry images. In *Proceedings of the Third International Symposium on Image Computing and Digital Medicine 2019*; pp. 102-106.
- [7] Sompawong N. Automated pap smear cervical cancer screening using deep learning. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2019*; pp. 7044-7048.
- [8] Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. *The Lancet* 2019; 393: 10167, 169-182.

- [9] Tao L, Amanguli A, Li F, Wang YH, Y, Yang, L, Mohemaiti M, Zhao J, Zou XG, Saimaiti A, Abudu M, Maimaiti M, Chen SY, Abudukelimu R, Maimati A, Li SG, Zhang W, Aizimu AA, Yang AQ, Wang J, Pang LJ, Cao YG, Gu WY, Zhang WJ. Cervical screening by Pap test and visual inspection enabling same-day biopsy in low-resource, high-risk communities. *Obstetrics & Gynecology* 2018; 132: 6, 1421-1429.
- [10] Chandran V, Sumithra MG, Karthick A, George T, Deivakani M, Elakkiya B, Subramaniam U, Manoharan S. Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images 2021; *BioMed Research International*. 5584004: 1-15.
- [11] Cibula D, Pötter R, Planchamp F, Avall-Lundqvist E, Fischerova D, Haie-Meder C, Köhler C, Landoni F, Lax S, Lindegaard JC, Mahantshetty U, Mathevet P, McCluggage WG, McCormack M, Naik R, Nout R, Pignata S, Ponce J, Querleu D, Raspagliesi F, Rodolakis A, Tamussino K, Wimberger P, Raspollini MR. The european society of gynaecological oncology/european society for radiotherapy and oncology/european society of pathology guidelines for the management of patients with cervical cancer. *Virchows Archiv* 2018; 472: 919-936.
- [12] Luo W. Predicting cervical cancer outcomes: statistics, images, and machine learning. *Frontiers in Artificial Intelligence* 2021; 4: 627369, 1-5.
- [13] Nayar S, Panicker JV, Nair JJ. Deep learning based model for multi-class classification of cervical cells using pap smear images. In 2022 IEEE 7th International Conference for Convergence in Technology (I2CT) 2022; pp. 1-6.
- [14] Ming Y, Dong X, Zhao J, Chen Z, Wang H, Wu N. Deep learning-based multimodal image analysis for cervical cancer detection. *Methods* 2022; 205: 46-52.
- [15] Ahishakiye E, Wario R, Mwangi W, Taremwa D. Prediction of cervical cancer basing on risk factors using ensemble learning,” In 2020 IST-Africa Conference (IST-Africa) 2020; pp. 1-12.
- [16] Saini SK, Bansal V, Kaur R, Juneja M. ColpoNet for automated cervical cancer screening using colposcopy images. *Machine Vision and Applications* 2020; 31: 1-15.
- [17] Priyanka BJ, Raju B. Machine learning approach for prediction of cervical cancer. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 2021; 12: 8, 3050-3058.
- [18] Mehmood M, Rizwan M, Gregus ml M, Abbas S. Machine learning assisted cervical cancer detection. *Frontiers in Public Health* 2021; 9: 788376, 1-14.
- [19] Alsmariy R, Healy G, Abdelhafez H. Predicting cervical cancer using machine learning methods. *International Journal of Advanced Computer Science and Applications* 2020; 11(7).173-184
- [20] Kalbhor M M., Shinde S V. Cervical cancer diagnosis using convolution neural network: feature learning and transfer learning approaches. *Soft Computing* 2023, 1-11.
- [21] Jantzen J, Dounias G. Analysis of pap-smear image data. The database can be downloaded at <https://mde-lab.aegean.gr/index.php/downloads/>. In *Proceedings of the Nature-Inspired Smart Information Systems 2nd Annual Symposium* 2006; 10: 1-11.
- [22] Mustafa WA, Kader MMMA. A review of histogram equalization techniques in image enhancement application. In *Journal of Physics: Conference Series* 2018; 1019: 1, pp. 1-7.

- [23] Dhal KG, Das A, Ray S, Gálvez J, Das S. Histogram equalization variants as optimization problems: a review. *Archives of Computational Methods in Engineering* 2021; 28: 3, 1471-1496.
- [24] Escorcía-Gutiérrez J, Mansour RF, Beleño K, Jiménez-Cabas J, Pérez M, Madera N, Velasquez K. Automated deep learning empowered breast cancer diagnosis using biomedical mammogram images. *Computers, Materials and Continua* 2022; 71: 3, 4221-4235.
- [25] Han X, Zhong Y, Cao L, Zhang L. Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sensing* 2017; 9: 848, 1-22.
- [26] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017; pp. 4700-4708.
- [27] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* 2017; 1704.04861.
- [28] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2018; pp. 4510-4520.
- [29] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2016; pp. 770- 778.
- [30] Chollet F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017; pp. 1251-1258.
- [31] Zheng Y, Yang C, Merkulov A. Breast cancer screening using convolutional neural network and follow-up digital mammography. In *Computational Imaging III* 2018; 10669: 1066905.
- [32] Cantor AB. Sample-size calculations for Cohen's kappa. *Psychological Methods* 1996; 1: 2, 150-153.
- [33] Çimen E. A transfer learning approach by using 2-d convolutional neural network features to detect unseen arrhythmia classes. *Eskişehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering* 2021; 22.1: 1-9.
- [34] Sun G, Li S, Cao Y, Lang F. Cervical cancer diagnosis based on random forest. *International Journal of Performability Engineering* 2017; 13: 4, 446-457.
- [35] Malli PK, Nandyal S. Machine learning technique for detection of cervical cancer using k-NN and artificial neural network. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 2017; 6: 4, 145-149.
- [36] Nguyen LD, Gao R, Lin D, Lin Z. Biomedical image classification based on a feature concatenation and ensemble of deep CNNs. *Journal of Ambient Intelligence and Humanized Computing* 2019. 1-13.
- [37] Lin H, Hu Y, Chen S, Yao J, Zhang L. Fine-grained classification of cervical cells using

- morphological and appearance based convolutional neural networks. *IEEE Access* 2019. 7: 71541-71549.
- [38] Khamparia A, Gupta D, de Albuquerque VHC, Sangaiah AK, Jhaveri RH. Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. *The Journal of Supercomputing* 2020; 76: 11, 8590-8608.
- [39] Ravindran K, Rajkumar S, Muthuvel K. An investigation on cervical cancer with image processing and hybrid classification. *International Journal of Performability Engineering* 2021; 17: 11, 918-925.
- [40] Lavanya Devi N, Thirumurugan P. Cervical cancer classification from pap smear images using modified fuzzy C means, PCA, and KNN. *IETE Journal of Research* 2022; 68: 3, 1591-1598.
- [41] Shinde S, Kalbhor M, Wajire P. DeepCyto: A hybrid framework for cervical cancer classification by using deep feature fusion of cytology images. *Math. Biosci. Eng* 2022; 19: 6415-6434.
- [42] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014; 1409.1556.
- [43] Tammina S. Transfer learning using VGG-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)* 2019; 9: 10, 143-150.