© Universiti Tun Hussein Onn Malaysia Publisher's Office

JSCDM

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/jscdm

Journal of Soft Computing and Data Mining

Deep Learning Approach for Predicting Prostate Cancer from MRI Images

Chin Wai Chun[1], Naeem Th. Yousir[2], Shaymaa Mohammed Abdulameer[2], Salama A. Mostafa[1*], Abdulkareem A. Hezam[1]

[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Batu Pahat, 86400, Johor, MALAYSIA

[2]College of Information Engineering,
Al-Nahrain University, 10072, Baghdad, IRAQ

*Corresponding Author

DOI: https://doi.org/10.30880/jscdm.2022.03.02.001
Received 06 June 2022; Accepted 01 August 2022; Available online 08 August 2022

Abstract: According to medical data, prostate cancer has been one of the most lethal malignancies in recent years. Early detection of prostate cancer significantly influences the tumor's treatability. Image analysis software that operates using a machine learning or deep learning algorithm is one of the techniques utilized to aid in the early and rapid identification of prostate cancer. This paper evaluates the performance of three deep learning Convolutional neural network (CNN) algorithms in detecting prostate cancer. Using Python, three deep learning models, ResNet50, InceptionV3, and VGG16, are subsequently created on the Kaggle platform. These three models have been applied to various medical image diagnostic problems and have won several contests. This study used 620 image samples from the Cancer Imaging Archive (TCIA) data source. Accuracy, f1 score, recall, and precision are used to evaluate the performance of the three models. The extracted test results indicate that the VGG16 achieves the highest level of accuracy at 95.56 percent, followed by the ResNet50 at 86.67 percent and the InceptionV3 at 85.56 percent.

Keywords: Deep learning, convolutional neural network, classification, prostate cancer, magnetic resonance imaging

1. Introduction

Prostate cancer is the second most often diagnosed cancer and the leading cause of cancer mortality among men worldwide, accounting for about 2 206 771 cases in 2020 [1]. According to U.S. statistics, 1 in 5 men will have prostate cancer in their lifetime, and half of the men may develop prostate cancer in their lifetime[1]. Men older than 50 will have an increased chance of developing prostate cancer. Prostate cancer diagnostics are performed via a prostate biopsy [2]. Patients with prostate cancer require tissue samples, which a physician will inspect under a microscope. The issue with prostate cancer biopsies is that it takes too long to get a result. The prostate cancer test will take between one and three days, although, for certain cases, the findings may take longer. In addition, PSA level testing is insufficient for diagnosing prostate cancer. Therefore, deep learning algorithms may be used to predict prostate cancer in order to minimize misdiagnoses and frequent mistakes made by physicians.

In addition, accuracy is poor; for instance, 30-40 percent of biopsies are false negatives. Fifteen percent of PSA tests may provide erroneous results, resulting in the patient receiving the incorrect medication. Using deep learning

**Corresponding author: salama@uthm.edu.my*

models with MRI, there are certain deep learning models with an accuracy of 90 percent for predicting prostate cancer [3]. Deep learning is a kind of artificial intelligence (AI) that can simulate human thought. Multi-Layer Perceptions (MLP), Convolutional Neural Networks (CNN), and Recurrent Neural Networks were the forms of deep learning (RNN). CNN models are used in medical therapy due to the fact that in deep learning, data may be filtered via numerous layers, and processing more data improves accuracy. In addition, deep learning's multiple-layer technique can categorize MRI pictures and data. In order to predict prostate cancer, this research will use three deep learning techniques: VGG16, Inception-V3, and ResNet. A few research publications demonstrate that deep learning is superior to humans. However, training the deep learning model requires a very large dataset. Deep learning may save more time and money when detecting prostate cancer than traditional methods [4].

This research aims to develop three deep learning models for predicting prostate cancer, namely VGG16, Inception-V3, and ResNet50. The second goal is to find the optimal parameters for deep learning models so that they can identify prostate cancer with a higher degree of precision. Last but not least, the performance of the generated prediction models will be evaluated using accuracy, precision, recall, and the F1 measure.

2. Related Work

Deep learning was used to construct a CNN model to predict prostate cancer in Liu [5] research. The development of the model was motivated by the fact that MRI is often regarded as the most reliable method for predicting prostate cancer; however, the accuracy also depends on the level of expertise and experience possessed by the attending physician. They used 10056 different diffusions weighted magnetic resonance imaging (DWI) pictures as datasets. Caffe was selected as the framework to use, and it is the only one that supports LMDB. Calculating mean files help improve testing and training states in terms of speed and accuracy. The images are arranged in training and testing sections, each with its own pictures. Seventy-five percent of the photos are utilized for training purposes, while the remaining twenty-five percent is used for testing purposes. For training purposes, there are 7575 DWI, of which 1539 are malignant, and 6036 are benign. During the testing, a total of 2481 pictures were utilized, 474 of which were malignant and 2007 of which were benign. The CNN used in this research has seven layers, and the size of its convolution kernels ranges from 5 pixels to 9 pixels to 11 pixels. The rate of accuracy is used as the measure in this investigation. When the model is in the training stage, its accuracy rate is 80.2 percent, but when it is in the testing state, this number is 78.2 percent, which is lower than the training state value. It is challenging to construct a deep learning model with a high accuracy rate with a limited dataset. This is due to the fact that deep learning is capable of extracting more and better characteristics if numerous MRIs are used. In the latter section of this study, there is some discussion of ways in which the situation may be improved. For instance, the use of a distinct kind of MRI could lead to distinguishable outcomes. The more layers in a CNN and the smaller the convolution kernels that are employed, the more varied the results that may be produced. In conclusion, the cross-validation approach may serve as a suitable substitute for the conventional technique.

Abdelmaksoud et al. [6] used both AlexNet and VGGNet to classify prostate cancer using transfer learning. In order to guarantee a high level of accuracy for models with a variety of values, a large number of trials are carried out. The datasets include 37 DWI with nine distinct b-values; of these, only 16 of the DWI are benign, while the remaining DWI are malignant. In order to get a result, DWI has to go through three processes: prostate segmentation, the computation of ADC maps, and the identification of ADC maps by using previously trained CNN models. There were 236 ADC maps that were benign and 234 ADC maps that were malignant. AlexNet anticipates receiving an input picture with dimensions of $227 \times 227 \times 3$, while VGGNet expected to get an image with dimensions of $224 \times 224 \times 3$. Seventy percent of the available ADC was used toward training, while the remaining twenty percent was utilized for testing. The metrics of accuracy, sensitivity, specificity, and precision were used in this research project. AlexNet achieved an accuracy of 89.2 ± 1.5 percent on average across all nine b-values, whereas VGGNet achieved an accuracy of 91.2 ± 1.3 percent on average across all nine b-values. According to the findings, using a CNN with a deeper layer may improve the accuracy of a prostate cancer diagnosis. There is still room for improvement in the findings by using a more in-depth CNN model such as ResNet. The findings demonstrate that the CNN models that are employed to identify prostate cancer are effective.

Salama et al. [7] experimented with VGG 16, ResNet50, and SVM. As a way to improve performance, SVM has been hybridized with VGG16 and ResNet50. The files include information on 1765 people, of whom 845 have prostate cancer, and 920 do not. A solution to the issue of inadequate DWI has been implemented, and that solution is data augmentation. For the purpose of this investigation, data augmentation was used, and new DWI pictures were produced by rotating the existing images at a different angle. In addition, the experiments were carried out twice, the first time with data augmentation and the second time without it. The augmentation was used to the rate of accuracy of the model by expanding the size of the input data by creating new data from the original data. The authors have used two splitting methods - train_test split (30-70%) and k-fold cross validation (1-5 folds) to accurately validate the results that were obtained later on. The acquired results showed that Resnet-50 and VGG16, when supplemented with additional data, achieve an accuracy rate of 94.74 percent 0.35 and 90.54 percent 0.22, respectively. VGG16 hybrid with SVM has an accuracy rate of 96.54 percent, whereas ResNet50 hybrid with SVM has an accuracy rate of 98.79 percent

According to the findings of this research, data augmentation could be used to enhance the number of samples, combat data shortage, and add variety to the dataset. The comparison and synopsis of these six investigations are shown in Table 1. According to the research findings, ResNet had the greatest performance, while VGGNet came in second.

Table 1 - The comparison and summary of the related studies

Citation	Dataset	Tool	accuracy(%)
[6]	470 sample -236 benign sample -234 malign sample	AlexNet	89.2 ± 1.5
		VGGNet	91.2 ± 1.3
[7]	1765 sample -845 are with PCa -920 are without PCa	Resnet-50	94.74%±0.35
		VGG16	90.54%±0.22
		VGG16 hybrid with SVM	96.54
		ResNet50 hybrid with SVM	98.79
[5]	10056 sample -2013 positive images -8043 negative images	CNN	80.2
		Decision Tree	99.85
		SVM Gaussian	100
		KNN-Cosine	99.85
		RUSBoost Tree	100

3. Methodology/Framework

3.1 Cross-Industry Standard Process for Data Mining (CRISP-DM)

The Cross-Industry Standard Process for Data Mining, often known as CRISP-DM, is an improvement upon earlier efforts to specify knowledge discovery methodologies [8], [9]. The CRISP-DM approach is widely used in the field of data mining. It has become an industry standard and can be described as a sequence of sequential phases that direct the deployment of data mining methods [10]. For the purpose of this study, the CRISP-DM will be implemented to extract the research outcomes. This approach consists of six steps or phases, the first of which is data understanding, as shown in Figure 1. Other phases are modeling, data preparation, business understanding, assessment, and implementation.

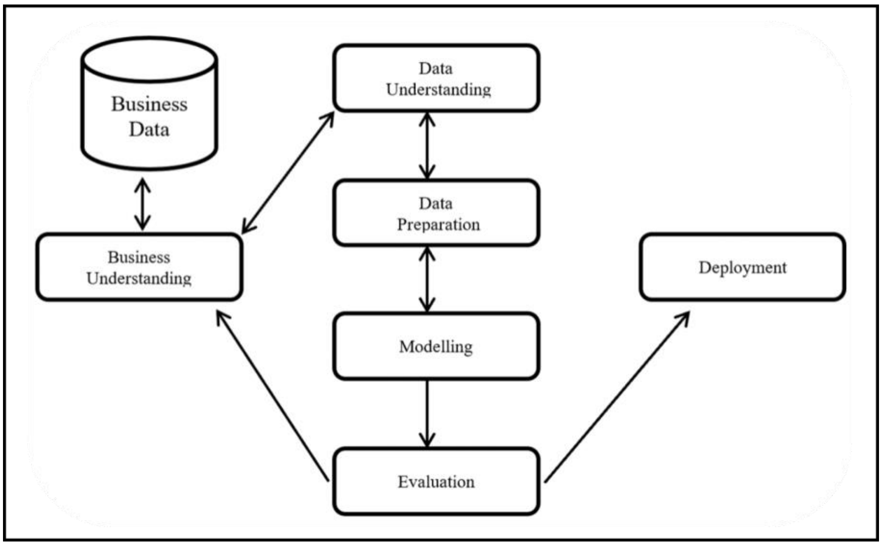


Fig. 1 - Phases of the CRISP-DM process model [10]

During the data understanding phase, the data will be gathered, studied, and understood to determine the reliability and quality of the data. It is very crucial to study the data thoroughly and understand its correlation and its suitability to be used for a deep learning project. In the data preparation phase, the preparation of the dataset takes place. In it, we preprocess the dataset, which includes tasks such as analyzing the data and transforming it into a more understood dataset so it won't raise issues or give false results when training the model. The modeling phase is mainly used for the classifiers' construction.

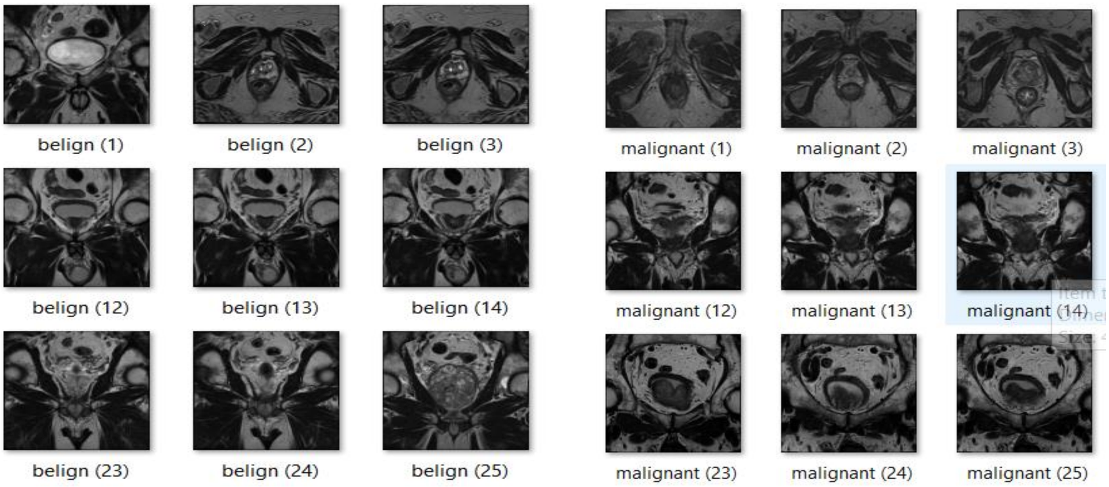
In this stage, we build the selected models, specify their parameters, and compile it. The last part is fitting the model and starting the model training stage with the preprocessed data. The evaluation phase is used to check the performance of the classifiers using the selected evaluation metrics: accuracy, loss, precision, recall, and f-measures. During the assessment phase, the performance of the deep learning model will be measured, compared, and recorded. A more in detail look into these phases can be observed in Table 2.

Table 2 - Summary and the outcome of each phase of CRISP-DM

Phases	Activity	Result outcomes
Data Understanding	Determine the objective of the project	• To increase the accuracy of deep learning models in detecting prostate cancer. • To develop prostate cancer prediction models by using VGG16, InceptionV3, and ResNet50 models. • To evaluate the performance of the developed prediction model using accuracy.
Data Understanding	Collect data	The data are collected from the cancer imaging archive;
	Explora data	The data have 620 samples of prostate cancer MRI. It is a 2D dimension with 256*256 pixels.
	Verify data quality	Accuracy
Data Preparation	Select Data	Give the reason for the data selected
	Clean Data	The clean samples that damaged data have been removed from it to get more accurate data.
Modeling	Select models	VGG16, InceptionV3 and ResNet50
	Test design	Each model will be trained for benign and malignant using 70% of the dataset. Then the models will be tested with the rest of the dataset.
	Build model	Select the parameter and describe the outcomes of the model;
Evaluation	Evaluation performance	The accuracy will be used to test the performance of the model; The result will be recorded and compared with each other.

3.2 Datasets

The dataset that we utilized for this investigation has a total of 620 magnetic resonance imaging (MRI) medical images. The dimensions of the images are 256 X 256 pixels. When classifying the MRIs, the amount of gray color in the image and the character or form of the gray region are also significant variables to consider. The images include prostate cancer (malignant) samples and samples with no prostate cancer (benign). Seventy percent of the images, or 440, will be used for training, and thirty percent, or 180, will be used for testing. During the training portion, there are 260 malignant MRIs and 180 benign MRIs, whereas, during the testing portion, there are 60 benign and 120 malignant MRIs (https://www.kaggle.com/datasets/waichunchin/mydataset4). Figure 2 displays some examples of the MRI datasets that were collected on prostate cancer patients.



(a) Benign samples (b) Malignant samples

Fig. 2 - MRI benign and malignant prostate cancer images

3.3 Algorithms

3.3.1 ResNet50

The acronym "ResNet-50" stands for "Residual Network," and it was first described in the work of Kaiming He [11]. The success of ResNet-50 in several contests is largely responsible for its widespread adoption. For instance, the ILSVRC classification competition and the COCO competition. In addition, there is an improvement of 28 percent compared to VGG, and it is simple and efficient to train numerous layers simultaneously. If you stack additional layers in the ResNet-50 model, you can enhance the performance and accuracy of the model. Figure 3 depicts the overall layout of the ResNet-50 system.

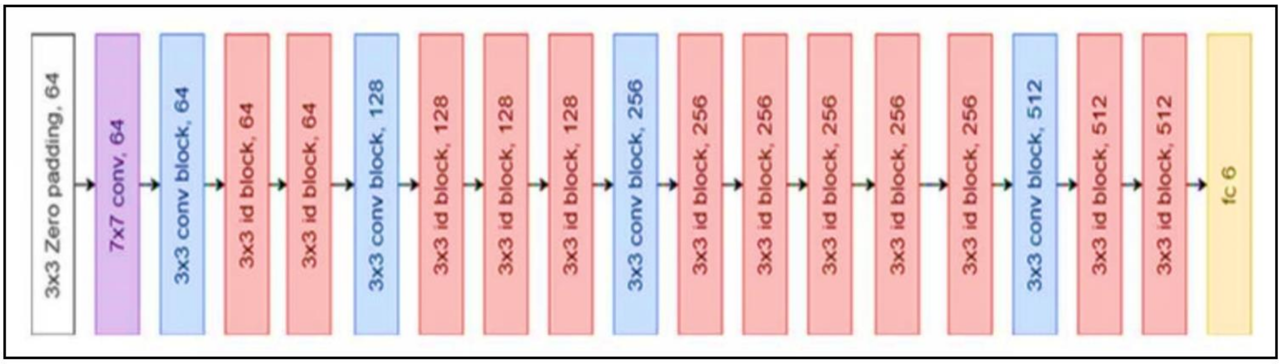


Fig. 3 - Architecture of ResNet50 [11]

3.3.2 VGG16

One of the most well-known CNN models is called VGG16. Simonyan et al. [12] were the ones who first brought it to light. Because it is simpler to implement, the VGG16 model is widely used for classifying pictures in various contexts. In addition, VGG16 is an enhancement over AlexNet since it includes 16 layers instead of only 12. The improvement brought about by VGG16's use of a 3X3 kernel size, in contrast to AlexNet's utilization of 11 and 5 for the first and second layers, respectively [13]. Figure 4 demonstrates the overall structure of the VGG-16 model.

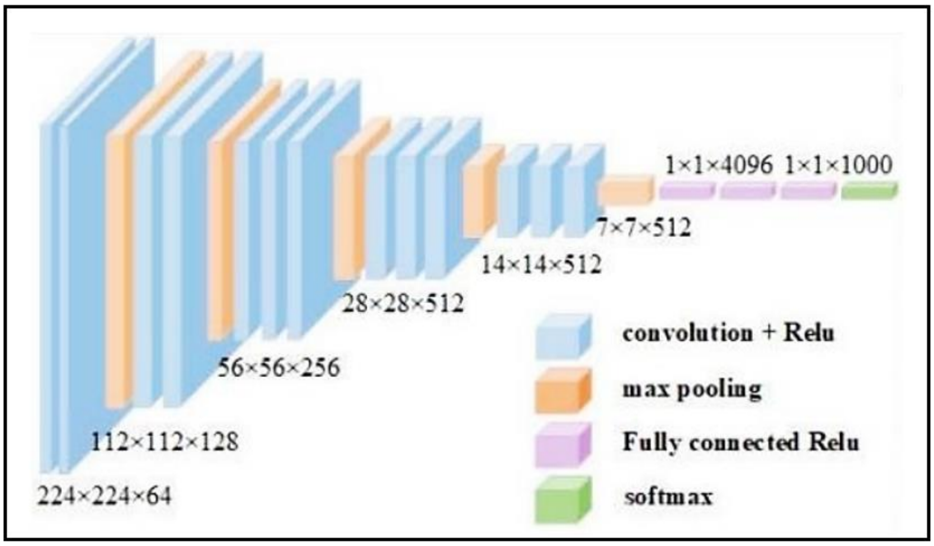


Fig. 4 - Architecture of VGG16 model [12], [13]

3.3.3 InceptionV3

InceptionV3 is a CNN model that is an improvement above InceptionV1 [14]. It is composed of 42 layers. InceptionV3 offers just a limited number of benefits. In contrast to the other models, InceptionV3 factorizes and uses a more compact convolution, and it has replaced the 5*5 convolutional layer with two 3*3 convolutional layers. The result is a decrease in the computing cost and the number of parameters. In addition to that, it will make use of a unique factorization into asymmetric convolutions. Convolutions of form 5*5 will be replaced, for instance, by convolutions of the kind 1*5 and 5*1. Furthermore, using auxiliary classifiers in inceptionV3 results in the network achieving greater accuracy and making it possible to become a deeper neutral network [15].

3.4 Evaluation Parameters

Accuracy and precision, as well as recall and F1 score, are the metrics that were used in this study to evaluate the three classifiers [16], [17]. At the study's conclusion, a comparison of how well ResNet50, VGG16, and Inception performed will occur. The following formulas show the mathematical representation of these four metrics:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$F1\ measure\ = 2\ X\ \frac{precision\ X\ recall}{precision + recall} \tag{4}$$

4. Results and Analysis

The reported result and analysis portion displayed a comparison of the performance of VGG-16, Inception-V3, and Resnet50. Accuracy, precision, recall, and F1 measure are the metrics that are being used for this project. One of these tools is called a confusion matrix. It displays the number of true positives, true negatives, false positives, and false negatives, making it simple to compute all relevant metrics. The confusion metrics of VGG16, InceptionV3, and ResNet 50 are shown in Table 3.

Table 3 - Confusion matrix of the three models

Model		True Positive	True Negative
VGG16	Predicted Positive	58	2
	Predicted Negative	6	114
InceptionV3	Predicted Positive	46	14
	Predicted Negative	12	108
ResNet50	Predicted Positive	51	9
	Predicted Negative	15	105

After we get the confusion matrix, we can compute all the metrics using the methods found in (1) through (4) through the use of true positive, true negative, false positive, and false negative results. During the training as well as the testing phase, accuracy was evaluated. Table 3 displays the models' performance outcomes throughout the project's testing phase. According to the findings, the performance of the VGG16 model, which has an accuracy of 95.56 percent, is the greatest. It is followed by the performance of the ResNet50 model, which has 86.67 percent accuracy. The model with the poorest performance was the InceptionV3 model, which has 85.56 percent accuracy.

Table 4 -The comparison results of the three models

Model	Accuracy	Precision	Recall	F1 Score
VGG16	0.9556	0.9667	0.9063	0.9355
InceptionV3	0.8556	0.7667	0.7931	0.7797
ResNet50	0.8667	0.8500	0.7727	0.8095

Figure 5 is a graph comparing the training accuracy results with the testing accuracy results throughout various epochs for the VGG16 model. The value of the number of epochs has been set to 50.

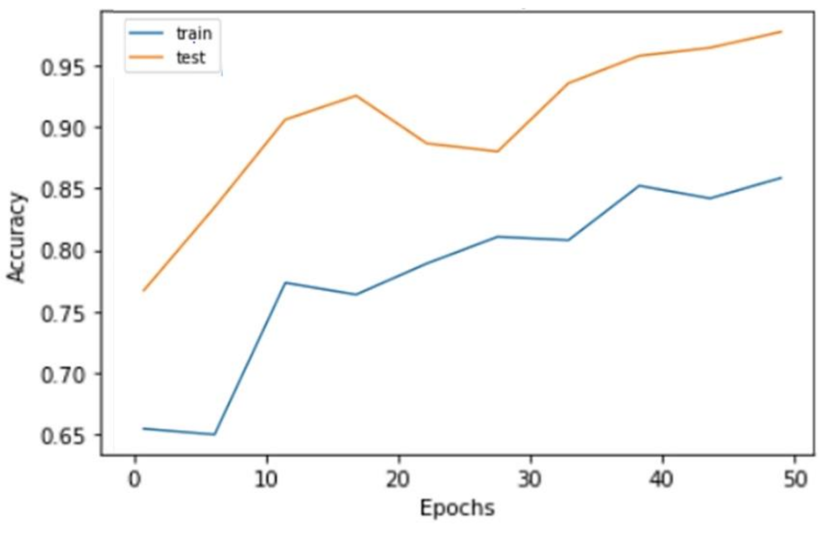


Fig. 5 - The training and testing results of the VGG16 model

Table 3 demonstrates that the performance of the models in the previous project produces better results than the current project for ResNet50 and InceptionV3. In contrast, the performance of VGG16 in the current project has a better performance score. However, the differences in the used datasets make it difficult to compare the previous work and this work [18], [19]. More specifically, this work uses limited prostate cancer images of 620 images which directly affect the training and performance of the deep learning models. There are many different approaches one may take to improve performance. For instance, modifying the epoch and batch size, increasing the dataset size, and increasing the percentage of the picture used for training from 70 percent to 80 percent of the total image used for training. The model that was used in the previous study was a more complicated and deeper network, which means it would need more training images to boost its accuracy. Despite this, the results demonstrate that the accuracy of the CNN model rose even when the sample size was 9000. In addition, CNN performs better than other models with larger sample size.

5. Conclusion

Prostate cancer is casual cancer among human males. Its test will take between one and three days, although, for some instances, the findings may take longer. This paper aims to automate prostate cancer diagnosis by using three deep learning models: ResNet50, InceptionV3, and VGG16. These models are tested using 620 prostate cancer images in which seventy percent of the prostate cancer images, or 440 of them, are used for training, and thirty percent, or 180, are used for testing. The performance of the training models is excellent, with every single outcome in training being at or above 50 percent. The VGG16 model provides the best overall performance of all the available options. The VGG16 achieves an accuracy of 95.56 percent, followed by the ResNet50 with an accuracy of 86.67 percent and the InceptionV3 achieves an accuracy of 85.56 percent. The number of epochs and the size of the dataset used in training a deep learning model may both be raised to provide better results. To enhance the performance of the model, new algorithms can be examined.

Acknowledgment

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, for supporting this work.

References

[1] Pollock, P. A., Ludgate, A., & Wassersug, R. J. (2015). In 2124, half of all men can count on developing prostate cancer. *Current oncology*, *22*(1), 10-12.

[2] Marks, L.S., Fradet, Y., Deras, I.L., Blase, A., Mathis, J., Aubin, S.M., Cancio, A.T., Desaulniers, M., Ellis, W.J., Rittenhouse, H. and Groskopf, J., 2007. PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. Urology, 69(3), pp.532-535.

[3] Erdem, E. and Bozkurt, F., 2021. A comparison of various supervised machine learning techniques for prostate cancer prediction. Avrupa Bilim ve Teknoloji Dergisi, (21), pp.610-620.

[4] Goldenberg, S.L., Nir, G. and Salcudean, S.E., 2019. A new era: artificial intelligence and machine learning in prostate cancer. Nature Reviews Urology, 16(7), pp.391-403.

[5] Liu, Y., & An, X. (2017, October). A classification model for the prostate cancer based on deep learning. In 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-6). IEEE.

[6] Abdelmaksoud, I.R., Shalaby, A., Mahmoud, A., Elmogy, M., Aboelfetouh, A., El-Ghar, A., El-Melegy, M., Alghamdi, N.S. and El-Baz, A., 2021. Precise Identification of Prostate Cancer from DWI Using Transfer Learning. Sensors, 21(11), p.3664.

[7] Salama, W.M. and Aly, M.H., 2021. Prostate cancer detection based on deep convolutional neural networks and support vector machines: a novel concern level analysis. Multimedia Tools and Applications, 80(16), pp.24995-25007.

[8] Adriaans, P., & Zantinge, D. (1996). Data mining addison wesley longman limited. Edinbourgh Gate, Harlow, CM20 2JE, England.

[9] Nodeh, M. J et al (2019, April). Analyzing and Processing of Supplier Database Based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) Algorithm. In The International Conference on Artificial Intelligence and Applied Mathematics in Engineering (pp. 544-558). Springer, Cham.

[10] Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1). London, UK: Springer-Verlag.

[11] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[12] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556..

[13] Pandiyan, Vigneashwara & Tjahjowidodo, Tegoeh & Caesarendra, Wahyu & Murugan, Pushparaja. (2019). “In-process virtual verification of weld seam removal in robotic abrasive belt grinding process using deep learning”. Robotics and Computer-Integrated Manufacturing. 57. 477–487. 10.1016/j.rcim.2019.01.006..

[14] Sam, S.M., Kamardin, K., Sjarif, N.N.A. and Mohamed, N., 2019. Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3. Procedia Computer Science, 161, pp.475-483.

[15] Ding, X., Guo, Y., Ding, G. and Han, J., 2019. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1911-1920).

[16] Obaid, O. I., Mohammed, M. A., Salman, A. O., Mostafa, S. A., & Elngar, A. A. (2022). Comparing the Performance of Pre-trained Deep Learning Models in Object Detection and Recognition. Journal of Information Technology Management, 14(4), 40-56.

[17] Chicho, B. T., & Sallow, A. B. (2021). A Comprehensive Survey of Deep Learning Models Based on Keras Framework. *Journal of Soft Computing and Data Mining*, *2*(2), 49-62.

[18] Abdulkareem, K. H., Mostafa, S. A., Al-Qudsy, Z. N., Mohammed, M. A., Al-Waisy, A. S., Kadry, S., ... & Nam, Y. (2022). Automated System for Identifying COVID-19 Infections in Computed Tomography Images Using Deep Learning Models. *Journal of Healthcare Engineering*, *2022*.

[19] Abdullah, S. M. S., & Abdulazeez, A. M. (2021). Facial expression recognition based on deep learning convolution neural network: A review. *Journal of Soft Computing and Data Mining*, *2*(1), 53-65.