Colorectal Cancer Diagnosis with Deep Learning Models

Merve Esra TAŞCI
Software Engineering Department, Sabahattin Zaim University
Istanbul, 34303, Turkey

Zahra ELMI
Computer Engineering Department, Sabahattin Zaim University
Istanbul, 34303, Turkey

Ömer Faruk ALBAYRAK
Computer Engineering Department, Sabahattin Zaim University
Istanbul, 34303, Turkey

Mustafa TOKAT
Computer Engineering Department, Sabahattin Zaim University
Istanbul, 34303, Turkey

ABSTRACT

The third most common disease in the world, colorectal cancer, frequently has the highest death rate. Surgery is a viable treatment option, but after five years, thirty to forty percent of patients have recurrence. Many people who have effectively treated their colorectal cancer also develop metastatic illness. Early detection is crucial since colorectal cancer has a high fatality rate. Deep learning techniques make colorectal cancer screening timelier and more cost-effective by enabling early and quicker identification of the disease. A collection of cell pictures was employed in the study to detect colorectal cancer. To demonstrate the capability of deep learning approaches, we used Convolutional Neural Networks (CNN), AlexNet, VGG-16, ResNet models and our proposed model as Hybrid CNN-LSTM. The accuracy and loss rates provided by the propos models were compared. The highest accuracy rate performance was observed with from the Hybrid CNN-LSTM model. The highest loss rate performance was observed with from the CNN model.

Keywords: Deep Learning colorectal cancer, CNN, Hybrid CNN-LSTM, ResNet, AlexNet, VGG16

1. INTRODUCTION

Colorectal cancer is one of the three most common malignancies diagnosed worldwide. It has been found to be among the most common causes of cancer-related mortality in another investigation. When cancer is detected at a specific stage, it can be treated surgically; however, it is known that thirty to forty percent of individuals who have the operation develop a recurrence of cancer. Therefore, detecting and initiating treatment for cancer before reaching the surgical stage is crucial [8].

In general, colorectal polyps may indicate the potential for colorectal cancer. The possibility of these polyps developing into cancer may take several years. Early detection of colorectal polyps allows for the use of essential treatments to stop them from developing into cancer. Colonoscopy is the most often used screening procedure for colorectal polyps. In 2012, guidelines for the surveillance of colorectal cancer after colonoscopy screening were established by the US Multisociety Task Force on Colorectal Cancer. These suggestions, which place a high priority on risk evaluation and guidance for follow-up, are based on the histological analysis of polyps discovered during the standard colonoscopy. Consequently, the detection and histological an essential part of colorectal cancer screening is the characterization of colorectal polyps to distinguish high-risk from low-risk polyps. This classification affects the likelihood of evolving colorectal cancer and increasing the amount of polyps in the future, as well as when to have follow-up colonoscopies [9].

Recent research on histology whole-slide image classification and segmentation tasks has shown that the deep learning approach outperforms traditional image processing methods. A platform for developing and applying computational models for illness diagnosis, patient care, and histopathological analysis of microscopic images has been established in the field of digital pathology as a result of the recent surge in the use of high-throughput tissue banks, whole-slide digital scanners, and the preservation of digital histological studies.

This study proposes deep learning methodology for computerized image analysis in light of recent

ISBN: 978-1-950492-79-4
ISSN: 2771-0947

https://doi.org/10.54808/WMSCI2024.01.92

developments and the urgent need for computational tools to support pathologists in the histopathological characterization and detection of colorectal polyps for more accurate and effective colorectal cancer screening [9]. These deep learning methodologies are Hybrid CNN-LSTM, CNN, ResNet, AlexNet, VGG16. Hybrid CNN-LSTM model is proposed for our study.

The study is set up these sections: Comparable traditional research is discussed in Section 2. The dataset, the picture classification procedure and the techniques used are all covered in Section 3, Materials and Methods. Section 4 presents a comparison of the models' accuracy rates. A summary of the study's objectives and the model with the best performance is provided in the last section.

2. RELATED WORKS

Numerous studies have been conducted on the detection of colorectal cancer. To identify patients with varying treatment responses and prognoses, Yamashita et al. have deemed Microsatellite Instability (MSI) identification in colorectal cancer to be crucial for clinical decision-making. They developed a method that uses pictures (WSI) immediately stained with hematoxylin and eosin (H&E) for MSI estimation using deep learning based and automated MSI prediction. Yamashita and colleagues used 100 randomly selected H&E-stained whole-slide images (WSIs) from 343 patients who had colorectal cancer resection at Stanford University Medical Center to create the deep learning model known as MSINet. Every picture was scanned with a magnification of 40×. They evaluated the model's performance to that of five gastrointestinal pathologists using a subset of 40 magnification word-wise inversions (WSIs) from an external dataset, which was randomly selected with 20 cases of Microsatellite Stability (MSS) and 20 cases of Microsatellite Instability (MSI) [2].

According to Kather and colleagues, almost all patients with colorectal cancer (CRC) have access to hematoxylin-eosin (HE) stained tissue slides. Quantitative data from these images, which are rarely utilized, can be objectively extracted to determine prognostic factors. In this work, they attempted to find out if CNNs might be used to extract prognostic factors—that is, whether or not a person has cancer—from these pictures. Based only on histology images, they have concluded that a CNN is capable of assessing the microenvironment of a human tumor and predicting prognosis [1].

Tsai and Tao [3] examined the CNN algorithm's performance with the AlexNet, SqueezeNet, VGGNet, GoogleNet and ResNet parameters. They discovered that the maximum accuracy rate was offered by the ResNet settings.

Li and colleagues in [5] claimed that liver metastasis (CRLM) is a symptom of colorectal cancer and early identification of CRLM is essential for an early cancer diagnosis. They employed computer tomography-derived contrast-enhanced computed tomography (CECT) scans in their investigation. They developed a dataset of CECT scans and applied cutting-edge deep learning methods to assess the most effective way to predict CRLM. According to experimental data, high-accuracy predictions were produced using a multi-plane architecture dubbed MPBD-LSTM, which was built on 3D bi-directional LSTM, an AUC (Area Under Curve) of 0.79 was attained.

Amirkhan et al. suggested that colorectal cancer could be accurately predicted using Electronic Medical Record (EMR) data. They were trained on EMR data to find out if LSTM and GRU networks could provide reliable colorectal cancer prediction models by learning temporal patterns from EMR data. The experimental results yielded an AUC rate of 0.811. [6].

Masum et al. employed a database containing details about 4336 patients who had colorectal cancer surgery, including age, length of operation, demographics, stoma status, complications, death rate and readmission. For predicting mortality, a variety of models were used, such as the Bidirectional Long Short-Term Memory (BI-LSTM) algorithms, Random Forest (RF), K-Nearest Neighbors (KNN), Multi-Layer Perceptron, Support Vector Machine (SVM) were used. According to this study, BI-LSTM algorithm surpassed previous models with the 0.841 accuracy.

3. MATERIAL AND METHOD

Material

In this study, the Colorectal Histology MNIST dataset [1] obtained from the Kaggle website was utilized. This dataset was generated from Kather and his colleagues [12] for the multiclass tissue classification problem. The textures in histological pictures of human colorectal cancer are represented by this data set. This dataset has included five thousand histological pictures showing eight distinct tissue types in human colorectal cancer. Images in the dataset of primary human colorectal adenocarcinomas embedded in paraffin (basic tumors) came from Heidelberg University's Institute of Pathology in Germany. Using Aperio ScanScope (Aperio/Leica biosystems), every RGB image with a pixel size of 0.495 µm was scanned. There was a 20x magnification set. Ten images belonging to eight textures are shown in Figure 1. The dataset used in our study is for colorectal cancer diagnosis.

[1] https://www.kaggle.com/datasets/kmader/colorectal-histology-mnist

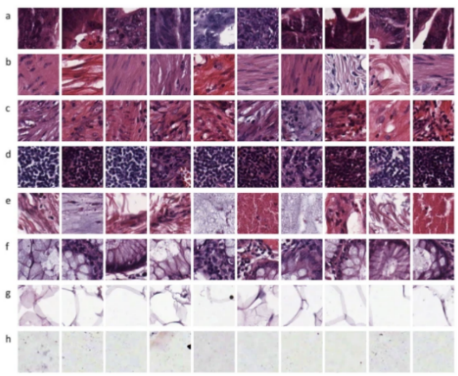
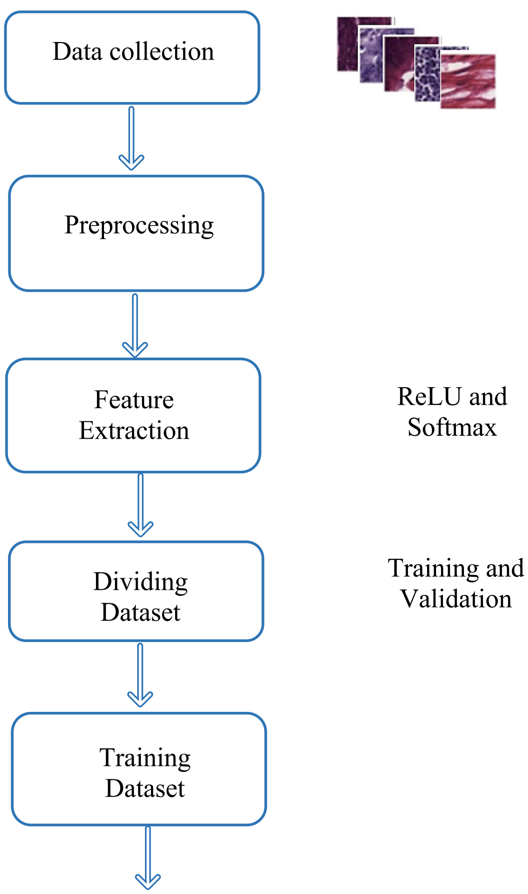


Figure 1. Images from MNIST dataset [12].

Method

CNN and Hybrid CNN-LSTM models were created and applied to the dataset for the diagnosis of colorectal cancer. AlexNet, VGG-16 and ResNet were used as parameters in the CNN model. The accuracy and loss rates obtained were compared with each other. Image classification method is illustrated in Figure 2.



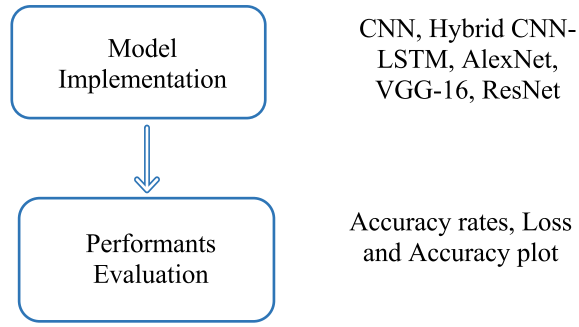


Figure 2. Image classification process

In our study we first had get colorectal cancer dataset from Kaggle. The dataset consists of 8 classes, with each class containing 625 images, totaling 5000 images. Out of these images, 4000 have been used for training and 1000 for validation. For CNN model, Sparse Categorical Crossentropy has been used as the loss function and Adam optimizer as a stochastic gradient descent with lr = 0.001. ReLU and Softmax activation functions have been applied in both algorithms. The CNN data has been trained for 128 epochs.

In our proposed model CNN-LSTM, units was developed to extract and classify complex features from temporal data sequences. The model utilizes the TimeDistributed wrapper to apply CNN layers to images at each time step in data sequences, such as time-dependent image series, and then feeds the extracted features into an LSTM layer. This approach enhances the ability to learn both spatial and temporal contexts. Data augmentation techniques were employed to improve the model's generalization capability, and various callback functions were optimized during the training process. Like CNN model, in the Hybrid CNN-LSTM, Sparse Categorical CrossEntropy has been used as the loss function and Adam optimizer as a stochastic gradient descent with lr = 0.001. Leaky-ReLU (alpha: 0.3) was used as the activation function to avoid saturation. The Hybrid CNN-LSTM model has been trained for 72 epochs and 128 batch sizes with fine-tuning.

In the ResNet50 model, we used Adam and Sparse Categorical Cross Entropy as an optimization algorithm and the loss function and a learning rate of 0.0001 following the guidance. The data has been trained for 100 epochs.

For VGG-16 model; out of these images, 4500 have been used for training and 500 for validation. Categorical Cross entropy has been utilized as the loss function and the Adam optimization algorithm with a learning rate of 0.001 is applied. The data has been trained for 20 epochs. For AlexNet model, 4500 have been used for training and 500 for validation. Sparse Categorical Crossentropy has been employed as the loss function and the Adam optimization algorithm with a learning rate of 0.001 is used. The data has been trained for 20 epochs. Similarly, in CNN and ResNet algorithms, ReLU and Softmax activation functions have been applied.

CNN

CNN, provide a few benefits for image processing and recognition. Specifically, these algorithms can use customized convolutional layers to efficiently extract features from images. They use parameters taken from features to learn complicated models. They can achieve great performance with few labeled data when used with data augmentation techniques. They can operate flexibly with inputs of varying sizes because of their sensitivity to input dimensions. CNNs use a modeling methodology that is more similar to human vision, producing useful results for practical uses. [11].

AlexNet

AlexNet was developed by Krizhevsky and colleagues. AlexNet has more significant characteristics than other deep learning methods. Each layer has numerous additional filters that can improve features and lower noise. A pooling layer, which can extract only the most important features and reduce the quantity of features, comes after each stacked convolution layer. Using Relu as the activation function instead of more biologically inspired functions like tanh, logistic, arctan, or sigmoid decreased the chance of the gradient vanishing. Based on the three points mentioned above, AlexNet performs five times better than other deep architectures. While certain deep architectures required specialized hardware, AlexNet can perform effectively with GPU and hardware constraints [13].

VGG-16

Karen Simonyan and Andrew Zisserman introduced the VGG-16 architecture for the first time in 2014. Karen and Andrew used only 3x3 convolutional layers to create a simpler 16-layer network. Thirteen convolutional layers, two fully connected layers and one SoftMax classifier make up the architecture of the VGG-16 model [14].

ResNet

ResNet50 is the shortened name for 50-layer residual networks. Both are very similar to VGG-16 ResNet50 has an additional identity mapping function compared to VGG-16. ResNet predicts the delta required to advance to the next layer. ResNet allows the gradient to pass through an alternative shortcut path, which covers the problem of the gradient disappearing.

If a CNN weight layer is not needed, the model can bypass it thanks to the identity mapping in ResNet. By doing this, overfitting to the training set is avoided. ResNet50 comprises fifty layers [10].

Hybrid CNN-LSTM

The LSTM architecture was developed to address the issues of short-term memory and gradient vanishing present in the RNN structure. In Figure 3, the architecture of the LSTM is depicted.

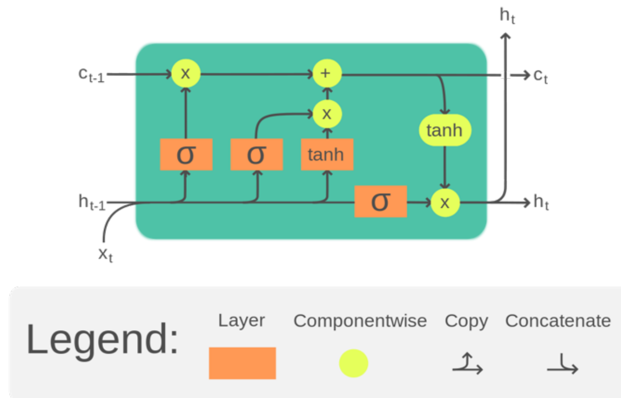


Figure 3. Architecture of the LSTM model [15].

The LSTM architecture consists of the memory state (cell state), forget gate, input gate, and output gate. The cell state is a structure that provides memory capability to the LSTM architecture, allowing meaningful information to be carried along a pathway for use when necessary. The forget gate decides which information to retain and which to discard. At the input gate, after applying sigmoid and tanh functions to previous and existing information, a decision is made on whether to update the memory state. After applying sigmoid and tanh operations, information resulting in 0 is considered unimportant, while information resulting in 1 is considered significant. The output gate determines what output to produce based on the input and cell memory. The information from the previous cell and the input information are passed through a sigmoid function. The information in the state cell undergoes a tanh function. The two results are multiplied to decide which information will be the input for the next cell [16].

In our study, a Hybrid CNN-LSTM model has been proposed to achieve higher performance. This model is created by using CNN and LSTM models together. The output obtained from the CNN model is used as input in the LSTM model.

4. RESULT AND DISCUSSION

In our study, the MNIST dataset was subjected to models from CNN, Hybrid CNN-LSTM, AlexNet, VGG-16 and ResNet. The accuracy and loss rates that were acquired were compared with each other.
Table 1 shows the accuracy rates obtained after applying CNN, Hybrid CNN-LSTM, ResNet, AlexNet and VGG-16 models. The highest accuracy rate is achieved by the Hybrid CNN-LSTM with 0.9240. Following the Hybrid CNN-LSTM, CNN achieved a rate of 0.9020, ResNet

attained an 0.88 accuracy rate, VGG-16 achieved rate of 0.7356 and AlexNet achieved rate of 0.7280.

Table 1. Accuracy rates of Hybrid CNN-LSTM, CNN, AlexNet, ResNet, VGG-16.

	Hybrid CNN-LSTM	CNN	AlexNet	ResNet	VGG-16
Accuracy rates	0.9240	0.9020	0.7280	0.88	0.7356

Table 2 demonstrates the loss rates obtained for Hybrid CNN-LSTM, CNN, ResNet, AlexNet and VGG-16 models. The lowest loss rate is achieved by the CNN model with a rate of 0.0176. Following the CNN model, Hybrid CNN-LSTM achieved a rate of 0.2298. ResNet achieved a loss rate of 0.23, AlexNet achieved a rate of 0.6916 and VGG-16 achieved a rate of 1.3875.

Table 2. Loss rates of CNN, ResNet, AlexNet, VGG-16 and Hybrid CNN-LSTM.

	CNN	Hybrid CNN-LSTM	ResNet	AlexNet	VGG-16
Loss rates	0.0176	0.2298	0.23	0.6916	1.3875

Figures 4, 5, 6, 7 and 8 show the plots of accuracy and loss rates acquired for training and validation for the proposed models. The CNN model has learned effectively from the applied dataset, as seen by the validation and training rates. When examining the Hybrid CNN-LSTM graph, it can be said that there is good learning as the training and validation rates converge and progress in parallel after a while. Looking at the loss rates, initially, the validation loss diverges from the training loss and then they converge and progress in parallel. By examining accuracy and loss rates, it can be said that the Hybrid CNN-LSTM model demonstrates good performance.

In ResNet graph, it can be stated that this model did not perform very well. Also, the training and validation lines diverge after a certain point, indicating that the model, instead of learning well, started to memorize. This suggests that the model excelled in training data but did not perform well on the validation dataset. By investigating the accuracy rate, AlexNet made predictions at 0.7280. Although the rate is not very high, it is an average value. When examining the graphs, it is observed that at some points, the validation and training rates diverge from each other. Despite not being excellent, AlexNet has shown better performance compared to ResNet.

The graph of the VGG-16 model shows that, although the loss rate is considerable, the training and validation lines follow each other closely. Another crucial factor that affects understanding the model's outstanding efficiency is the loss rate. It is a 1.3875 loss rate. Based on this rate, the VGG-16 model has performed the worst out of all of the models on the used database.

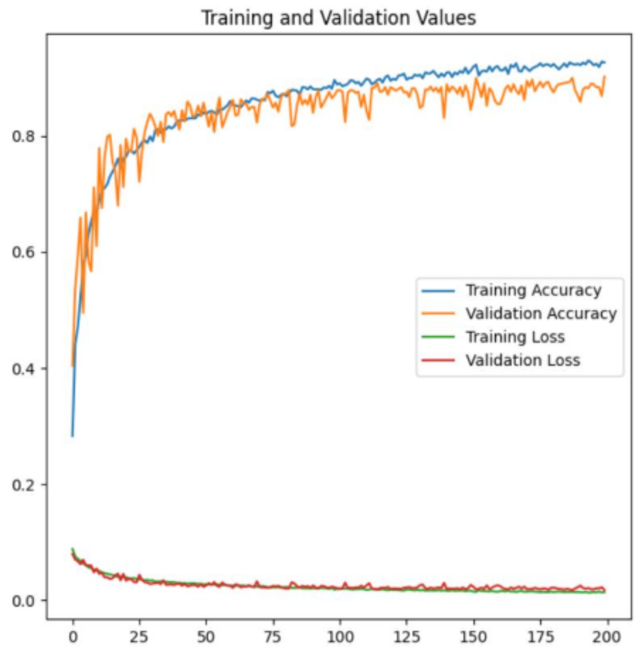


Figure 4. CNN accuracy and loss rates plot.

Figure 5. ResNet accuracy and loss rates plot.

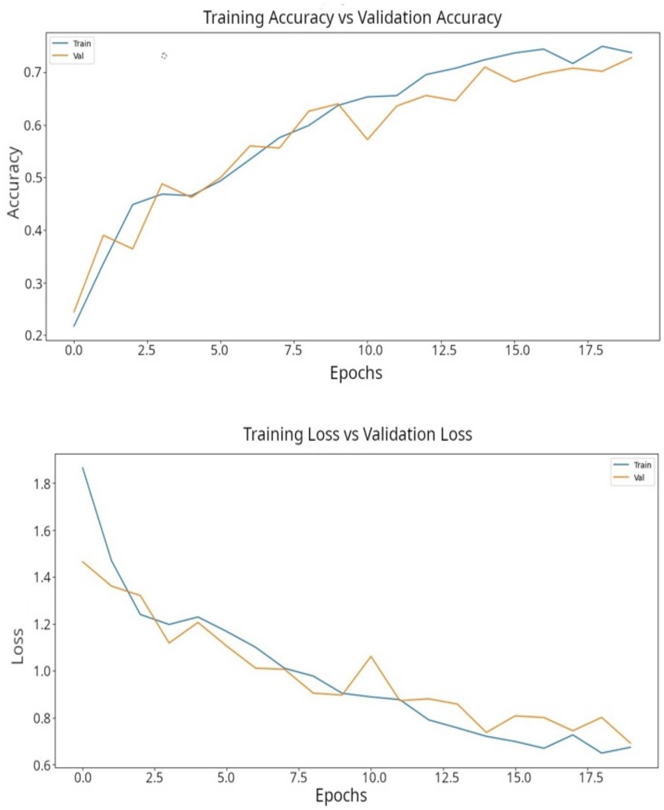


Figure 6. AlexNet accuracy and loss rates plot.

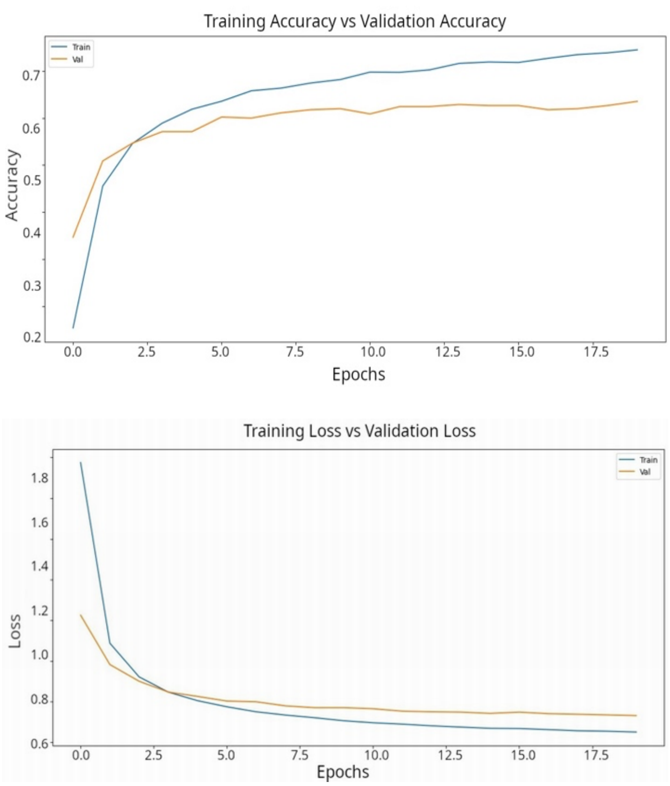


Figure 7. VGG-16 accuracy and loss rates plot.

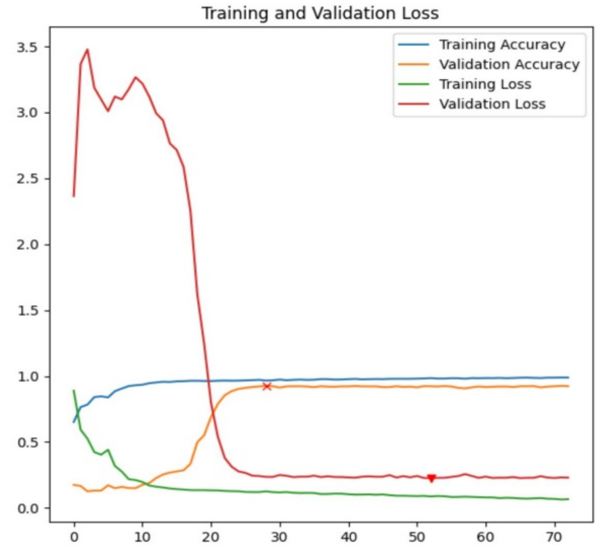


Figure 8. Hybrid CNN-LSTM accuracy and loss rates plot.

5. CONCLUSION

One of the most prevalent cancer types worldwide is colorectal cancer. Cancer can be surgically treated if it is found at a certain stage. In more severe situations, surgery is the only way to recuperate. It is estimated that thirty to forty percent of patients who have surgery may eventually get a cancer recurrence. It's critical to identify cancer before it reaches the surgical stage. Deep learning techniques make it possible to identify colorectal cancer earlier, which saves time and money. This study made use of the MNIST dataset. The accuracy rates of CNN, AlexNet, VGG-16, ResNet models and proposed model Hybrid CNN-LSTM were compared. The model with the best accuracy rate performance was Hybrid CNN-LSTM with 0.9240, following this CNN give 0.9020. The lowest loss rate was given by CNN model with 0.0176. We could conclude that our proposed model Hybrid CNN-LSTM and CNN model demonstrate the best performance in comparison with ResNet, AlexNet and VGG-16 models.

6. REFERENCES

[1] Kather, J. N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C. A. and Halama, N., **Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study.** PLOS Medicine, 16(1)

[2] Yamashita, R., Long, J., Longacre, T., Peng, L., Berry, G., Martin, B. and Shen, J., **Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study.** The Lancet Oncology, 22(1)

[3] Tsai, M.J., Tao, Y. H., **Deep Learning Techniques for the Classification of Colorectal Cancer Tissue.** Electronics 2021, 10, 1662

[4] Ho, D., Tan, I. B. H. and Motani, M., **Predictive models for colorectal cancer recurrence using multi-modal healthcare data**. Proceedings of the Conference on Health, Inference and Learning

[5] Li, X., Xiao, H., Weng, W, Xu, X. and Shi, Y., **MPBD-LSTM: A Predictive Model for Colorectal Liver Metastases Using Time Series Multi-phase Contrast-Enhanced CT Scans**, International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2023: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023 pp 379–388

[6] Amirkhan, R., Hoogendoorn, M., Numans, M. E. and Moons, L., **Using recurrent neural networks to predict colorectal cancer among patients**. IEEE Symposium Series on Computational Intelligence (SSCI)

[7] Masum, S., Hopgood, A., Stefan, S., Flashman, K. and Khan, J., **Data analytics and artifcial intelligence in predicting length of stay, readmission and mortality: a population-based study of surgical management of colorectal cancer**, Discover Oncology (2022) 13:11

[8] Ho, D., Chong, D. Q., Tay, B., Tan, I. B. and Motani, M. **Prognosticating Colorectal Cancer Recurrence using Machine Learning Techniques**, IEEE International Conference on E-Health Networking, Application & Services (HEALTHCOM).

[9] Korbar, B., Olofso, A.,M., Miraflor, A., P., Nicka, C., M., Suriawinata, M., A., Torresani, L., Suriawinata, A, A. and Hassanpour, S., **Deep Learning for Classification of Colorectal Polyps on Whole-slide Images**, Journal of Pathology Informatics, Volume 8, Issue 1, January–December 2017, 30

[10] Theckedath, D. and Sedamkar, R. R., **Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks**, SN Computer Science, 1(2)

[11] Simonyan, K. and Zisserman A., **VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION**, 3rd International Conference on Learning Representations (ICLR 2015) 1-14

[12] Kather, J., N., Weis, C., A., Bianconi, F., Melchers, S., M., Schad, L., R., Gaiser, T., Marx, A. and Zollner, F., **Multi-class texture analysis in colorectal cancer histology**, Scientific Reports (in press)

[13] Hosny, K. M., Kassem, M. A. And Fouad, M. M., **Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet,** Journal of Digital Imaging

[14] Tammina, S., **Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images**, International Journal of Scientific and Research Publications, Volume 9, Issue 10, October 2019

[15] Kara, A. (2019), **Global Solar Irradiance Time Series Prediction Using Long Short-Term Memory Network**. Gazi Üniversitesi, Fen Bilimleri Dergisi, Part C: Tasarım ve Teknoloji, 7(4): 882-892.

[16] Gavcar, E. ve Metin, H., M. (2021), **Hisse Senedi Değerlerinin Makine Öğrenimi (Derin Öğrenme) ile Tahmini**. Ekonomi ve Yönetim Araştırmaları Dergisi, 10(2).