Indonesian Journal of Data and Science (IJODAS) ISSN: 2715-9930
Vol 4, No 1, March 2023, pp. 25-30

Diagnosis of Hepatitis using Supervised Learning algorithm

Musa Dima Genemo [a,1]

[a] *Gumushane University, Bağlarbaşı, 29100 Gümüşhane Merkez/Gümüşhane, Turki*
musa.ju2002@gmail.com;

ARTICLE INFORMATION

Diterima : 10 – 02 – 2023
Direvisi : 26 – 03 – 2023
Diterbitkan : 31 – 03 – 2023

Kata Kunci:
Classification,
Feature Fusion,
Hepatitis,
Image Processing

ABSTRACT

Hepatitis is the most serious disease in developing countries. Therefore, early diagnosis is very important to obstacle the effect that can happen as a consequence of this disease. In this case, deep learning can solve the issue at an early stage. An innovative deep learning-based technique to identify hepatitis is presented in this study. In this study 45 layers, convolutional neural network (CNN) architecture connected with three fully connected layers is used in the proposed architecture. The two classes of collected hepatitis datasets are then used to train the suggested CNN model. The model achieved 0.934 classification accuracy. The proposed model was compared to the state of the art at the time. The outcome presented implies that the model's performance is remarkable.

I. Introduction

Hepatitis is the most serious diseases in developing country. This disease is known by affecting liver cells, and inflammatory cells in the liver tissue. It can be caused by alcohol, taking drugs without physician order, contamination with patient of the disease and lack of vaccination at early childhood. There are five (5) main hepatitis viruses, referred to as types A, B, C, D and E. These 5 types are of greatest concern because of the burden of illness and death they cause and the potential for outbreaks and epidemic spread. Among all hepatitis virus, Type B virus (HBV) is a serious disease caused by a virus that attacks the liver. Chronic hepatitis B virus (HBV) infection affects 240–300 million people worldwide [1,2,3]. Hepatitis B virus (HBV) infection is a serious global health problem despite the availability of prophylactic vaccine and antiviral treatments [4]. HBV causes a variety of liver diseases in humans, from acute to chronic diseases which may result in liver cirrhosis and hepatocellular carcinoma (HCC) [5,6]. In the Figure 1 you can see health and chronic hepatitis.

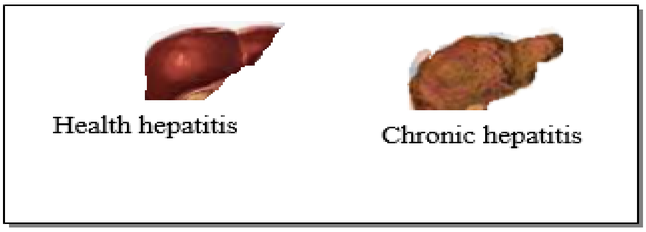

Figure 1. health and chronic hepatitis

The early detection of hepatitis can reduce the mortality rate. Automated Deep Learning models can be utilized for disease recognition at an initial stage because the manual identification of virus is time-consuming and less efficient. These systems are based on the supervised learning technique and predict diseases using handcrafted and deep learning (DL) features. The researchers have utilized point features [7], color features [8], texture features [9, 10], and convolutional neural network (CNN) feature [11-15] from pre-trained models. DL has become a crucial domain for the recognition and segmentation tasks to implement image processing techniques on medical images. DL is one of the robust machine learning (ML) tools used for the automated classification of medical lesions into their pertinent categories [16-18].

It can enhance the identification process of diseases and help health professionals [19]. The CNNs are the deep learning models utilized for feature derivation from gastric images. A CNN can undertake data in different forms, such as signals, images, videos, and multi-dimensional. A basic CNN architecture incorporates different layers, including convolutional, max pooling, fully connected (FC), and classification. The initial layers of CNN provide low-level features, but the deeper layers of the architecture provide high-level features. The convolutional layers perform convolution operations on all pixels of the image input data. The FC layers vectorize the extracted features for disease recognition [20]. In this work, a novel deep CNN-based technique is designed for the classification of hepatitis diseases.

The other parts of the paper are arranged as section 2 reviews the existing work; section 3 gives the proposed technique; section 4 demonstrates the results and discussion, and section 5 is the conclusion of the work.

The existing literature has shown promising performance for accurate disease detection as compared to manual diagnosis. Different machine learning techniques and algorithms have been used for the analysis of several disease like Gastric Infections (GI), Tumors, and Hepatitis. In this section some existing study and their result discussed.

A CNN model [24] was presented based on the spatial attention mechanism with encoder and decoder layers to classify GI diseases. In experiments, the model was validated by adopting a five-fold cross-validation approach. Experiments have shown that the calculated accuracy, precision, recall, and F1-score are 93.19%, 92.8%, 92.7%, and 92.8%, respectively.

In [24], the authors performed extensive experiments to obtain promising results for the classification of GI anomalies. They picked the five renowned pre-trained CNN architectures Xception, Inception-v3, MobileNet, ResNet, and VGG-16, and re-trained them on the Kvasir database. The classification accuracies obtained using Xception, Inception-v3, MobileNet, ResNet, and VGG-16 were 98.2%, 90%, 97.6%, 92.3%, and 98.3%, respectively. It shows that maximum accuracy 98.3% was observed on the VGG-16 model.

Another work [25] presented a framework for the classification of stomach lesions based on the DL features. The proposed framework used VGGNet and InceptionNet architectures for the computation of deep features. After extraction of feature sets from both deep networks, feature sets were concatenated and assigned to the multiple ML classifiers for disease identification. The highest disease recognition accuracy was achieved by the ML algorithm SVM, and the calculated accuracy rate was 98%. Sutton et al.

In [26], researchers took advantage of different hybrid features to categorize endoscopic images. For low-level feature extraction, a bottom-up multi-scale gradual integration (BGMI) technique was utilized, and the input was rectified by top-down modules for attention learning. To compute high-level features, the backbone stream was utilized.

II. Proposed Methodology

In this Section the proposed methodology is presented with detailed modeling and explanation. Figure 2 represent Architecture of the proposed model for the diagnosis of hepatitis diseases. The stapes explained in here is the development of new Model, and CNN architecture, feature extraction and feature fusion.

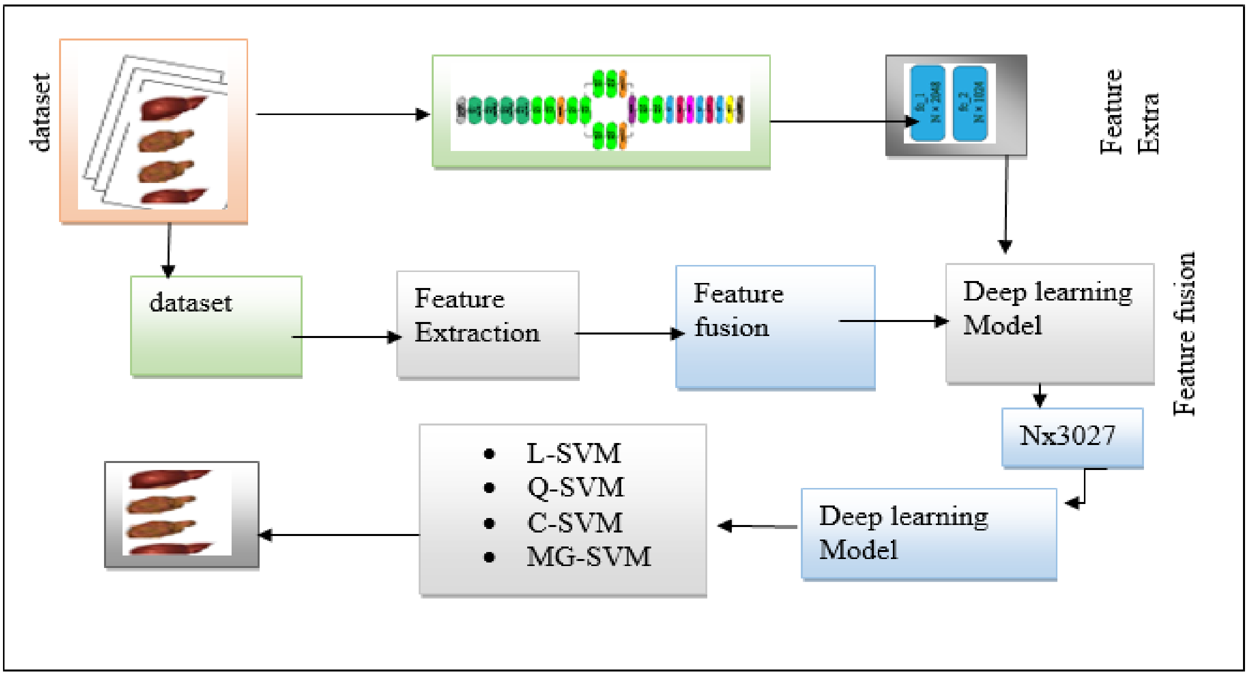

Figure 2. Architecture of the proposed model for the classification of hepatitis diseases

A novel CNN model is designed comprising 45 layers. The proposed architecture contains 14 convolutional layers, 16 ReLU layers, four average pooling layers, three batch normalization layers, a dropout layer, and three FC layers. An image input layer is used to input image data into a network. The input image size of image input layer is 256×256×3. Equation one represents convolutional operation as follows

$$O(i,j) = \sum_{b=0}^{c} \sum_{a=0}^{r} Filter(a,b).I(i-a,j-b) \tag{1}$$

Where O is output, C is column, R is row, and I is input.

RELU Activation function is applied on every element and negative values are replaced with zero. Then pooling is used to downsize the feature map. The mapping function h can be mathematically represented as:

$$h(m_i, W) = Wm_i \tag{2}$$

Where m denotes the input set, and W represents the weight matrix.

The final layer of a deep CNN architecture is the class output layer. In this layer, the cross-entropy loss is computed between mutually exclusive classes to perform classification tasks. In the Figure 3 you can see 45-Layered construction of the suggested CNN model

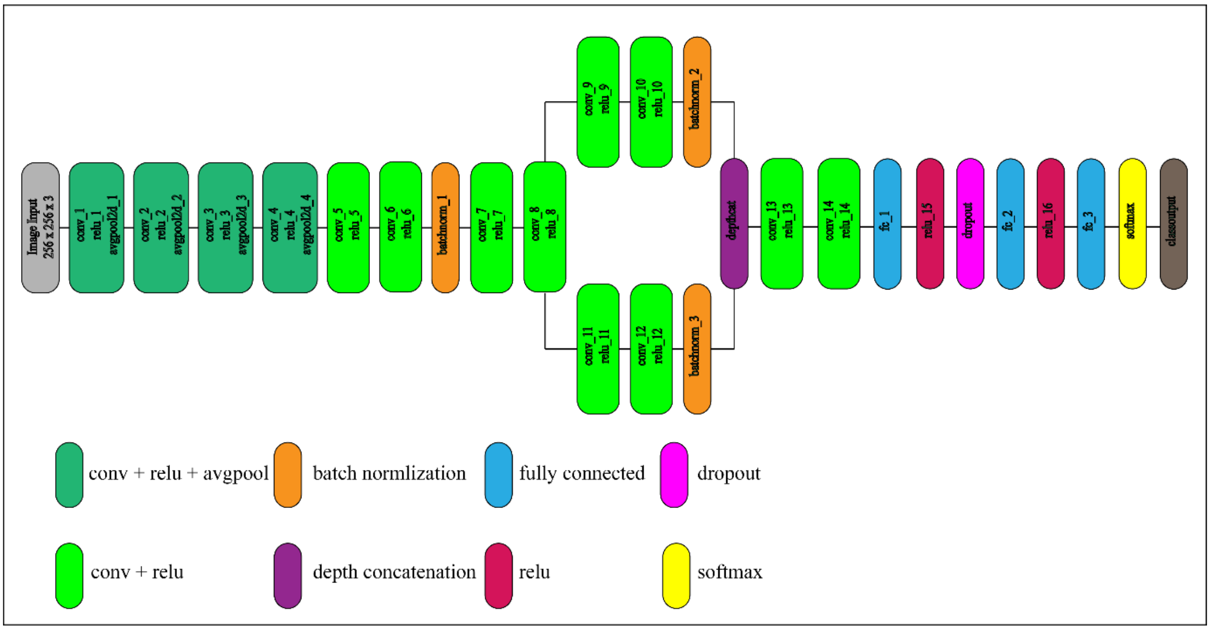

Figure 3. 45-Layered construction of the suggested CNN model

The model is trained on the dataset for feature learning. The training is performed in 6 epochs with learning rate of constant value 0.001as presented in Error! Reference source not found.. From the total data 75% of the data is set for training, and 25% of the data is for validation. After the training of the network, deep features from the two FC layers fc_1 and fc_2 are extracted for the classification of gastric infections. Both FC layers fc_1, and fc_2 produced $N \times 2048$, and $N \times 1024$ dimensionality features sets, respectively. The N is the total training images. The extracted feature sets $N \times 2048$, and $N \times 1024$ are combined into a single feature vector using a serial concatenation procedure. The final feature vector of size N×3072 is achieved after the concatenation of features. After all, the fused feature vector is given to the multiple ML predictors for the recognition of diseases. Mathematically, the fusion of features can be given as:

$$\psi(fused)_{N\times3072} = \psi(fc\_1)_{N\times2048} + \psi(fc\_2)_{N\times1024} \tag{3}$$

III. Result and Discussion

A. Result

In this section the procedure and complete description of the experiments taken is presented. The experiment is performed on data UCI Machine Learning Repository Hepatitis Dataset. In the experiments, two classes (chronic and health liver) are utilized to predict the result. The experiments are performed on defined ratios of data from both training and testing with a cross-validation approach of 10-folds.

The C-SVM classifier took 28.491 seconds for the training which is the lowest in contrast to the other predictors in the experiment. The additional performance measures such that precision, recall, and F1-score computed on the C-SVM classifier are 0.934, 0.924, and 0.919, respectively

Table 1. Results of 50:50 training and testing data for stomach disease classification

Classifier	Acc (%)	Prc (%)	Rc (%)	F1-S (%)	Time (Seconds)
L-SVM	0.890	0.901	0.890	0.920	39.902
Q-SVM	0.928	0.920	0.919	0.920	28.987
C-SVM	0.934	0.924	0.919	0.925	28.491
MG-SVM	0.909	0.914	0.909	0.916	30.521
ESD	0.916	0.920	0.899	0.931	86.55
Neural Network	0.919	0.919	0.918	0.920	38.942

As we may noticed from Table 1 the best computational time is 28.987 seconds observed on the neural network classifier.

B. Discussion

In this study we presented a new model for hepatitis diesis diagnosis and classification. During training the model, deep features are extracted and fused into a single feature set and extensive experiment is performed on datasets. The minimum computational time recorded during training and the maximum accuracy achieved is 28.9491 and 0.934 respectively. The highest accuracy is scored on C-SVM, and the best and worst computational time is 39.902 seconds, and 28.491 respectively.

Table 2. Maximum accuracy calculated on different training and testing ratios

Classifier	Accuracy	Computational Time
L-SVM	0.890	39.902
Q-SVM	0.928	28.987
C-SVM	0.934	28.491
ESD	0.927	29.987
Q-SVM	0.919	29.987
C-SVM	0.919	29.987

We can observe from the Table 2 that, there is no big difference in classifiers accuracy except computational time. Moreover, the neural network classifier gives superior performance in terms of recognition time.

IV. Conclusion

Medical image processing is having been an important area of research in recent year. In this study a new deep learning model is introduced to detect hepatitis virus diagnosis. This work is encompassed to diagnosis HBV patients image using a new CNN model on two classes of dataset (affected and health liver). During training the model feature extraction, feature fusion is used to classify the dataset as affected and health liver. After computing experimental results, the maximum prediction rate of 0.934 is achieved which shows that the proposed CNN architecture obtained promising performance. In the future, by increasing number of varies datasets, we work on improving classification results and simulation of the system.

References

[1] Hepatitis B Foundation. Hepatitis B facts and figures. Accessed October 20, 2021.

[2] Tang LSY, Covert E, Wilson E, Kottilil S. *Chronic hepatitis B infection: a review*. JAMA. 2018;319:1802-1813.

[3] European Association for the Study of the Liver. *EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection*. J Hepatol. 2017;67:370-398.

[4] Zoulim F, Lebosse F, Levrero M. *Current treatments for chronic hepatitis B virus infections*. Curr Opin Virol. 2016;18:109-116.

[5] Ott JJ, Stevens GA, Groeger J, Wiersma ST. *Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity*. Vaccine. 2012;30(12):2212-2219.

[6] Mohd Hanafiah K, Groeger J, Flaxman AD, Wiersma ST. *Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence*. Hepatology. 2013;57(4):1333-1342.

[7] Rusnak R, Ganem D. Sequences *5′ to the polyadenylation signal mediate differential poly(a) site use in hepatitis B viruses*. Genes Dev. 1990;4(5):764-776

[8]. Tuba, E., M. Tuba, and R. Jovanovic. *An algorithm for automated segmentation for bleeding detection in endoscopic images*. in *2017 International Joint Conference on Neural Networks (IJCNN)*. 2017. IEEE.

[9] Majid, A., et al., *Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection.* 2020. **83**(5): p. 562-576.

[10] Li, B. and M.Q.-H.J.I.T.o.I.T.i.B. Meng, *Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection.* 2012. **16**(3): p. 323-329.

[11] Ali, H., et al., *Computer assisted gastric abnormalities detection using hybrid texture descriptors for chromoendoscopy images.* Computer methods and programs in biomedicine, 2018. **157**: p. 39-47.

[12] Khan, M.A., et al., *StomachNet: optimal deep learning features fusion for stomach abnormalities classification.* 2020. **8**: p. 197969-197981.

[13] Amin, J., et al., *Liver Tumor Localization Based on YOLOv3 and 3D-Semantic Segmentation Using Deep Neural Networks.* Diagnostics, 2022. **12**(4): p. 823.

[14] Khan, M.A., et al., *Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine.* Computers & Electrical Engineering, 2021. **90**: p. 106960.

[15] Khan, M.A., et al., *An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection.* Neural Computing and Applications, 2020. **32**(20): p. 15929-15948.

[16] Khan, M.A., et al., *Developed Newton-Raphson based deep features selection framework for skin lesion recognition.* Pattern Recognition Letters, 2020. **129**: p. 293-303.

[17] Amin, J., et al., *An integrated design for classification and localization of diabetic foot ulcer based on CNN and YOLOv2-DFU models.* IEEE Access, 2020. **8**: p. 228586-228597.

[18] Amin, J., et al., *An integrated design for classification and localization of diabetic foot ulcer based on CNN and YOLOv2-DFU models.* IEEE Access, 2020. **8**: p. 228586-228597.

[19] Sharif, M., et al., *Recognition of different types of leukocytes using YOLOv2 and optimized bag-of-features.* IEEE Access, 2020. **8**: p. 167448-167459.

[20] Khan, M.A., et al. *Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification*. in *2019 international conference on computer and information sciences (ICCIS)*. 2019. IEEE.

[21] Khan, M.A., et al., *Stomach deformities recognition using rank-based deep features selection.* 2019. **43**(12): p. 1-15.

[22] Khan, M.A., et al., *Multiclass Stomach Diseases Classification Using Deep Learning Features Optimization.* 2021.

[23] Lonseko, Z.M., et al., *Gastrointestinal disease classification in endoscopic images using attention-guided convolutional neural networks*. 2021. 11(23): p. 11136.

[24] Dheir, I.M. and S.S.J.I.J.o.A.E.R. Abu-Naser, *Classification of Anomalies in Gastrointestinal Tract Using Deep Learning.* 2022. 6(3).

[25] Haile, M.B., et al., *Detection and classification of gastrointestinal disease using convolutional neural network and SVM.* 2022. 9(1): p. 2084878.

[26] Li, S., et al., *Effective high-to-low-level feature aggregation network for endoscopic image classification.* 2022: p. 1-9.