

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Classification of Hepatocellular Carcinoma Using Machine Learning

*Lekshmi Kalinathan, Deepika Sivasankaran,  
Janet Reshma Jeyasingh, Amritha Sennappa Sudharsan  
and Hareni Marimuthu*

## Abstract

Hepatocellular Carcinoma (HCC) proves to be challenging for detection and classification of its stages mainly due to the lack of disparity between cancerous and non cancerous cells. This work focuses on detecting hepatic cancer stages from histopathology data using machine learning techniques. It aims to develop a prototype which helps the pathologists to deliver a report in a quick manner and detect the stage of the cancer cell. Hence we propose a system to identify and classify HCC based on the features obtained by deep learning using pre-trained models such as VGG-16, ResNet-50, DenseNet-121, InceptionV3, InceptionResNet50 and Xception followed by machine learning using support vector machine (SVM) to learn from these features. The accuracy obtained using the system comprised of DenseNet-121 for feature extraction and SVM for classification gives 82% accuracy.

**Keywords:** Hepatocellular Carcinoma, Feature extraction, Convolution Neural Networks, Prognosis, Machine Learning

## 1. Introduction

The existing work on Hepatic tumor is concerned with clinical data acquired through blood samples, urine samples and serum test, and non-invasive images like CT, MRI, PET and SPECT. The manual identification of cancer from microscopic biopsy images is subjective in nature and may vary from expert to expert depending on their expertise and other factors which include lack of specific and accurate quantitative measures to classify the biopsy images as normal or cancerous one. Stains such as Hematoxylin and Eosin (H and E stain) are used for better emphasis of the nuclei of liver cells. Based on the amount of stain absorbed by the nuclei, it can be classified into various types since nuclei size increases with the stages of cancer. The stain can also be accumulated on the tissues causing ambiguity to the pathologist. Such ambiguity in the images can be overlooked by an individual. Color normalization is done to highlight the nuclei for visually better features.

Normalization techniques discussed in the study [1] where the images are classified by their colors using K Means Clustering and JSEG segmentation In this method, the nuclei get segmented as a separate segment. Then it is passed onto the SVM classifier. This technique enables effective segmentation of colored images. Similarly JSEG segmentation technique has two phases: color quantization and spatial segmentation [2]. Color quantization is based on peer group filtering(PGF) and vector quantization to reduce the number of colors in the images. For addressing the drawbacks of JSEG method, contrast map and improved contrast map were obtained. This technique saw a significant improvement in detecting more homogeneous regions than that of JSEG method. Due to the inherent difficulty involved in obtaining liver cell images from the biopsies, Liangqun et al. proposed to use neural networks for feature extraction and SVM for classification [3]. This method aims at providing better efficiency from less number of images.

The findings of the study [4] demonstrated the capability of Convolutional Neural Network (CNN) to recognize distinct features that can detect tumor masses in a histopathological liver tissue image. The author proposed to implement the CNN model for segmentation and classification of different stages of HCC. However, the major drawback of using CNNs for the feature extraction process is that these models need large amounts of data to process. This is a huge challenge for the biomedical field as it is pragmatically difficult to have access to massive data. Moreover, feature learning is pertinent on the size, shape and degree of annotation of images which are not uniform across datasets.

Chen et al. developed a deep convolutional neural network to classify the lung tumor stage and predict the most commonly mutated genes in lung cancer tissue cells [5]. Ehteshami et al. also produced a promising result for the classification of breast tumors using deep learning techniques [6]. The author developed an algorithm to differentiate stroma invasive cancer and stroma from benign biopsies However, the deep learning models were applied to non solid tumors. Thus, it remains uncertain if they can produce the same accuracy when applied to solid tumors.

2. Proposed methodology

The workflow contains 4 modules as follows:

- 1. Data collection
- 2. Color normalization
- 3. Creation of a classifier

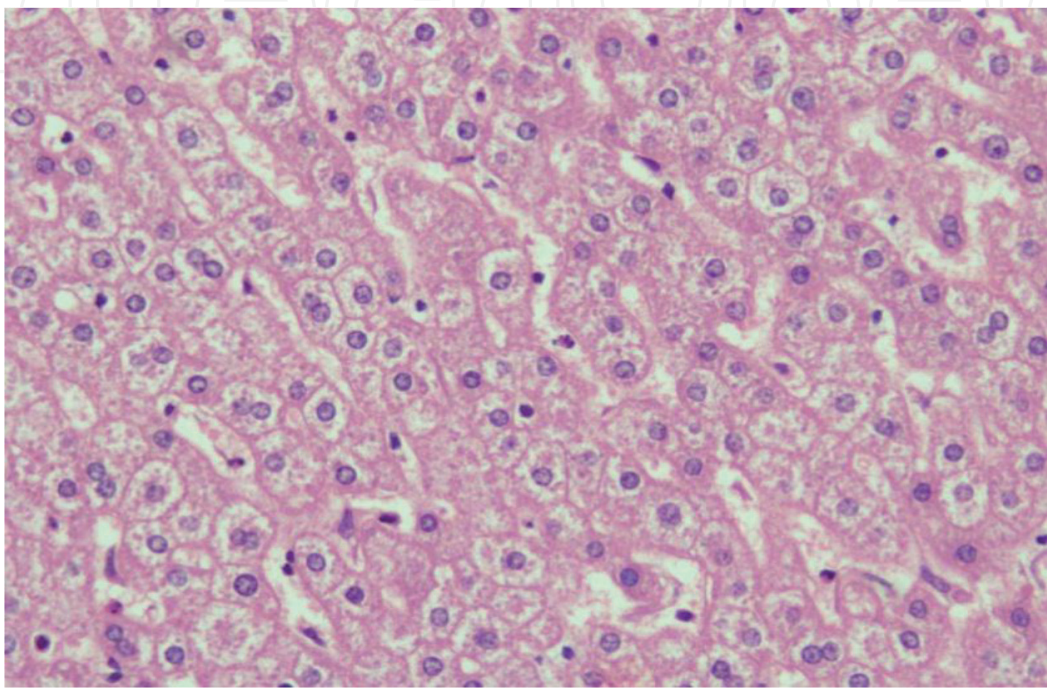
Cancer type	Images
Non cancerous	232
Well-differentiated carcinoma	148
Moderately differentiated carcinoma	81
Poorly differentiated	189
TOTAL	687

Table 1.  
HCC dataset split-up.

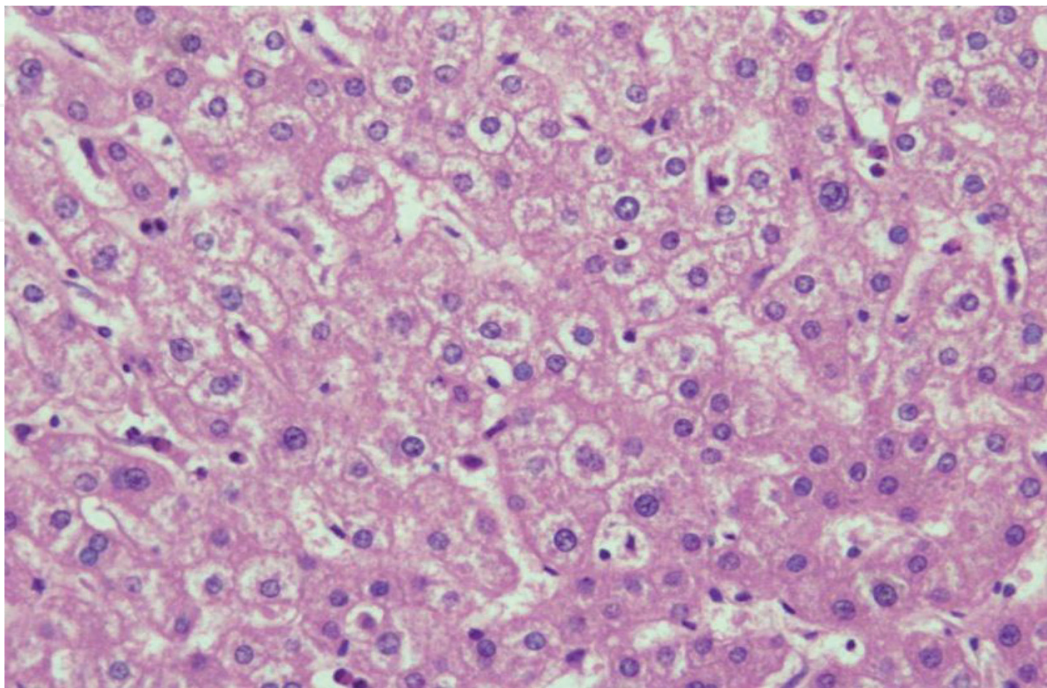
2.1 Data collection

The first phase involved collection of data from Dataset collected from Global Hospital, Perumbakkam, Chennai. In a span of 3 weeks, images were collected from the biopsies of 3 patients. The three types of cancerous images obtained during the data collection phase are well-differentiated, moderately differentiated and poorly differentiated. The total number of images collected is 687 whose split up is given in **Table 1**.

Below are some images from the dataset collected, **Figures 1–4**.

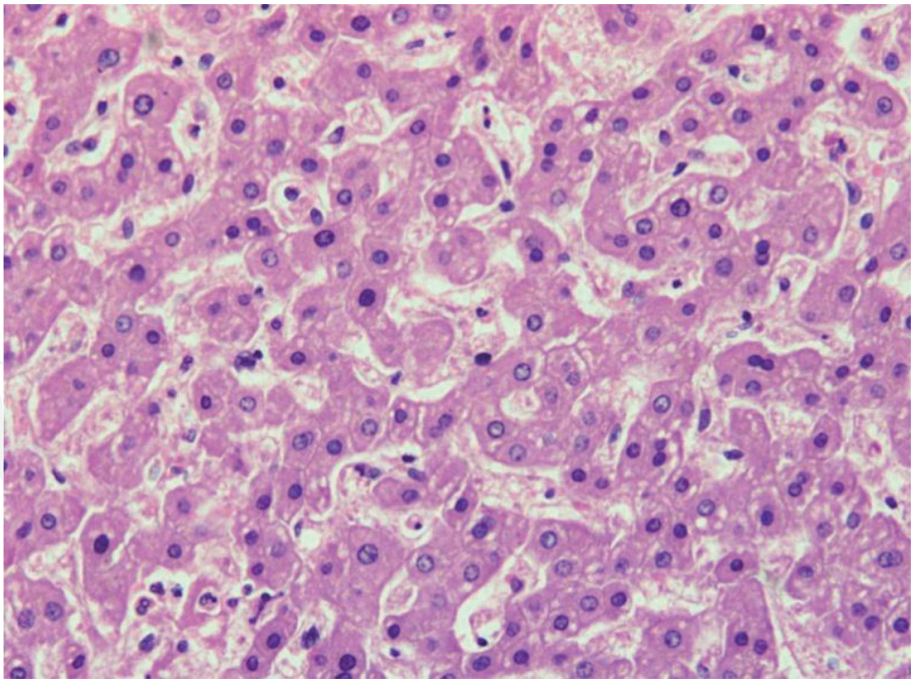


**Figure 1.**  
*Non cancerous image.*

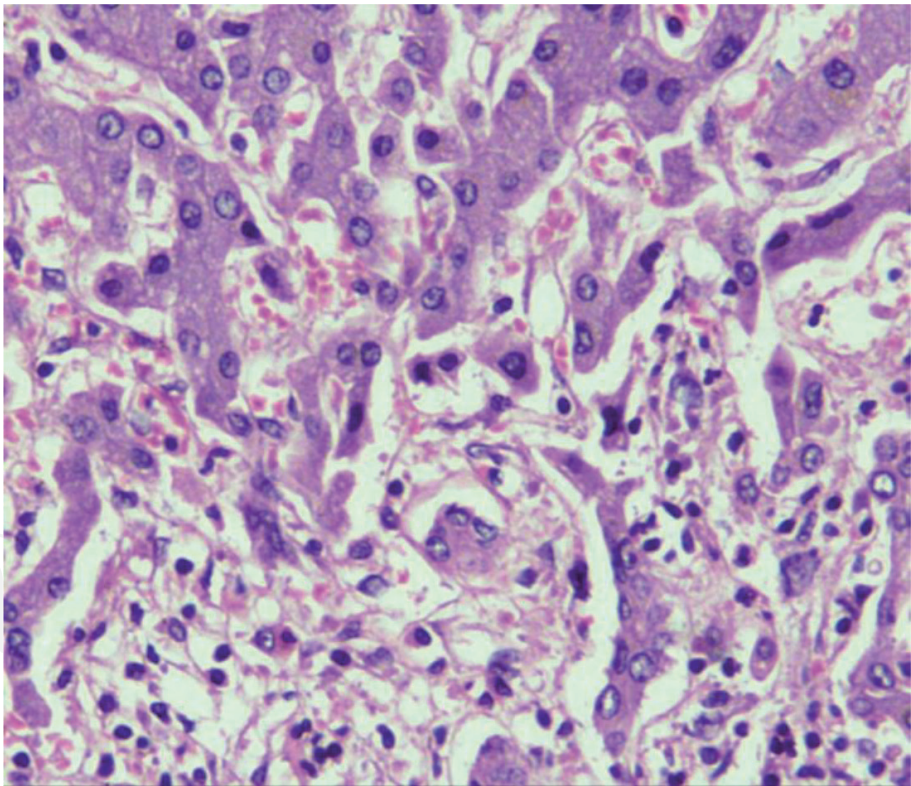


**Figure 2.**  
*Well differentiated cancer.*





**Figure 3.**  
*Moderately differentiated cancer.*

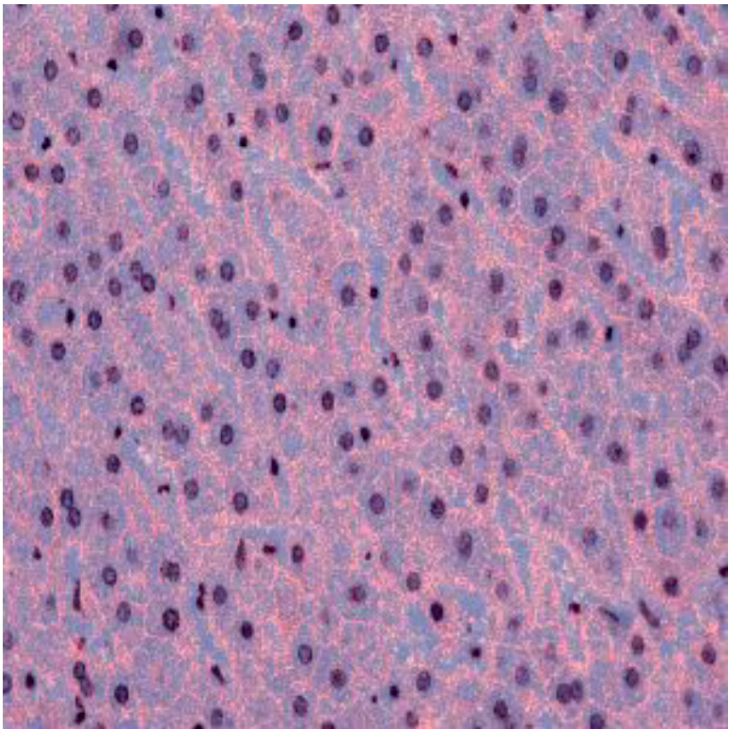


**Figure 4.**  
*Poorly differentiated cancer.*

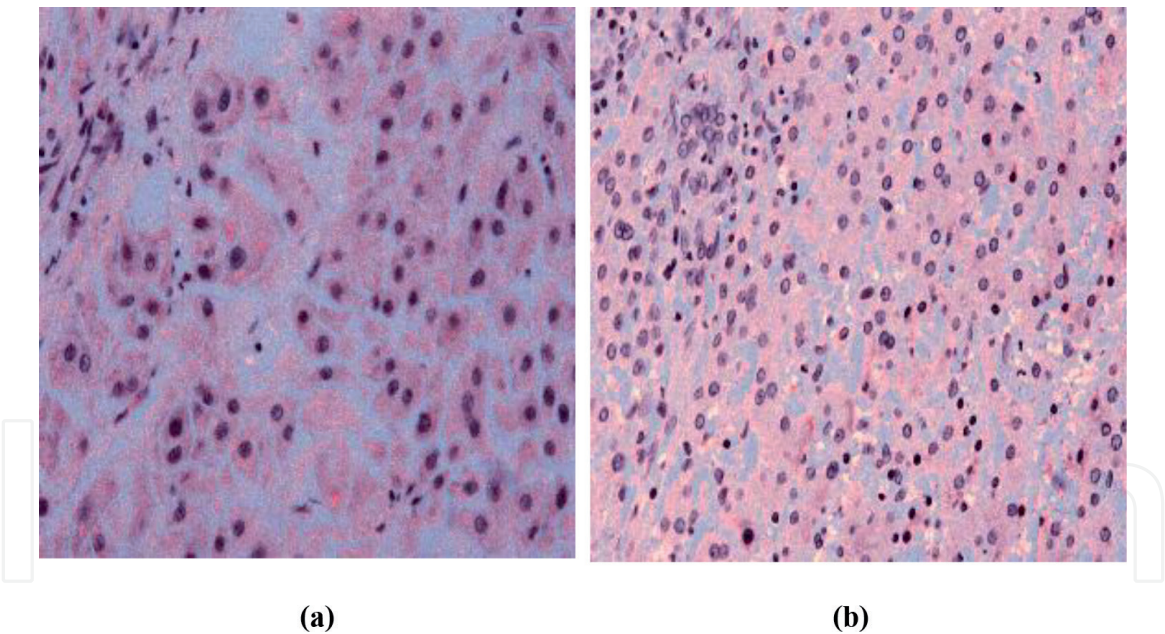
**2.2 Color normalization**

The features of the nuclei include the texture, size and roundness. Applying a stain on these biopsies cause the nuclei to be highlighted due to absorption of the stain. The color difference between the nuclei and the tissues may be visually comparable or less different. Hence, color normalization is done to highlight the nuclei. Highlighting the nuclei makes it easier to extract the features from them.





**Figure 5.**  
*Normalized non cancerous image.*



**Figure 6.**  
*(a), (b) normalized cancerous images.*

The normalization method [3] is exclusive to H and E stain. Normalized images are shown below (Figures 5 and 6).

### 2.3 Creation of a classification system

Using convolution neural networks (CNN) can be less efficient in creating a classifier system mainly due to its requirement of a large dataset to learn from. Using CNN is not a very practical approach as it may not be feasible to collect a dataset containing large numbers of images. Thus an alternative method is proposed

where features are extracted from the images using unsupervised deep learning and then a supervised machine learning classifier is used to learn from those features for classification. The advantage of this method is the elimination of overfitting of the class with majority data and the system can work fairly well with less number of images. Using a support vector machine (SVM) the classifier is built and pretrained models such as VGG-16, ResNet50, DenseNet –121, DenseNet –169, DenseNet-201, InceptionV3, InceptionResNet50 and Xception.

3. Performance analysis

To select the best feature extractor from all the pretrained models, metrics such as F1- score and accuracy are considered. Higher accuracies may not be the most efficient and reliable metric always. Hence, F1-score is also considered as it shows individual class performance and is useful when the dataset is highly imbalanced. **Table 2** shows the overall accuracies obtained when all the pretrained models are used.

From **Table 2**, it is found that performance of DenseNet is better than the other deep learning architectures. The performance of the variants of DenseNet is given in **Table 3**. Here it is observed that with the increase in the number of layers of DenseNet from 121 to 201, there is a degradation in the accuracy. Hence, the F1 score is also affected.

S. no	Model	Accuracy (%)
1	Xception	72
2	VGG16	78
3	ResNet50	80
4	InceptionV3	74
4	InceptionResNetV2	45
5	DenseNet	85

**Table 2.**  
*Performance of various pretrained models with SVM.*

S. no	Model	Accuracy (%)
1	DenseNet –121	82
2	DenseNet –169	84
3	DenseNet –201	81

**Table 3.**  
*Performance of DenseNet variants.*

S. no	Classifier	Accuracy (%)
1	DenseNet –121 + SVM	82
2	DenseNet –121 + Naive Bayes	70
3	DenseNet –121 + Decision Tree	61

**Table 4.**  
*Performance of DenseNet –121 with the classifiers.*

Class	Precision	Recall	F1-score	Support
Non cancerous	0.79	0.83	0.81	69
Well-differentiated cancer	0.86	0.81	0.83	37
Moderately differentiated cancer	0.58	0.67	0.62	21
Poorly differentiated cancer	0.97	0.88	0.93	42
Accuracy			0.82	169
Macro average	0.80	0.80	0.80	169
Weighted Average	0.83	0.82	0.82	169

**Table 5.**  
*Performance of DenseNet –121 with SVM.*

The final pretrained architecture selected for feature extraction is DenseNet –121 to be combined with the machine learning classifiers. Supervised algorithms such as decision tree, SVM, Naive bayes were taken into consideration to find the optimal classifier. The results of the feature extractor and classifier are given in **Table 4**. From **Table 4**, SVM is chosen to be the optimal classifier that works best with DenseNet –121 feature extractor.

DenseNet-121 is chosen due to high f1-score in spite of having less accuracy than DenseNet-169. Performance analysis of DenseNet-121 is given in **Table 5**.

4. Conclusions and future work

From the results obtained, it is observed that this method can provide better accuracy although the dataset is highly imbalanced and when there is a deficit in the dataset. Using convolution neural networks (CNN) can underperform when the dataset is imbalanced and it requires an extensive dataset to learn from. Improvements can be made by obtaining more data. Procuring more images from biopsies and medical data will help improve the system’s efficiency and this can be extended as a separate component for the microscope.

Author details

Lekshmi Kalinathan\*, Deepika Sivasankaran, Janet Reshma Jeyasingh,  
Amritha Sennappa Sudharsan and Hareni Marimuthu  
Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar  
College of Engineering, Chennai, India

\*Address all correspondence to: [lekshmik@ssn.edu.in](mailto:lekshmik@ssn.edu.in)

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.





## References

[1] Himanshu Yadav, Prateek Bansalt and Ramesh Kumar Sunkaria, Color Dependent K-Means Clustering for Color Image Segmentation of Colored Medical Images, 1st International Conference on Next Generation Computing Technologies (NGCT-2015), 2015

[2] Yu-Chou Chang, Dah-Jye Lee, Yong-Gang Wang, Color-Texture Segmentation of Medical Images Based on Local Contrast Information, IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), ( 2007).

[3] Liangqun Lu and Bernie J. Daigle, Jr. (2020). Prognostic analysis of histopathological images using pre-trained convolutional neural networks: application to hepatocellular carcinoma, PeerJ, doi 10.7717/peerj.8668

[4] Azer, Samy A. (2019) “Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: A systematic review.” *World journal of gastrointestinal oncology* vol. 11,12: 1218-1230. doi:10.4251/wjgo.v11.i12.1218

[5] Chen, Mingyu & Zhang, Bin & Topatana, Win & Cao, Jiasheng & Zhu, Hepan & Juengpanich, Sarun & Mao, Qijiang & Yu, Hong & Cai, Xiujun. (2020). Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *npj Precision Oncology*. 4. 14. 10.1038/s41698-020-0120-3.

[6] Ehteshami Bejnordi, B. et al. (2018) Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod. Pathol.* 31, 1502-1512.