ISSN: 1672 - 6553

JOURNAL OF DYNAMICS AND CONTROL

VOLUME 9 ISSUE 3: 109 - 115

ENHANCED LIVER CANCER DETECTION USING HYBRID CNN-XGBOOST MODEL IN MACHINE LEARNING

Kalaiselvi N.[1], Sasikala R.[2]

Student[1], Assistant professor[2],
Department of Computer Science and Engineering,
KIT – Kalaignarkarunanidhi Institute of Technology, Coimbatore, India

DOI: https://doi.org/10.71058/jodac.v9i3008

ENHANCED LIVER CANCER DETECTION USING HYBRID CNN-XGBOOST MODEL IN MACHINE LEARNING

KALAISELVI N.[1], SASIKALA R.[2]

Student[1], Assistant professor[2]
Department of Computer Science and Engineering
KIT – Kalaignarkarunanidhi Institute of Technology, Coimbatore, India

kalaiselvi2613@gmail.com[1], kitsasicse@gmail.com2

Abstract: *Since liver cancer is such an aggressive illness, improving patient outcomes requires early identification. This paper presents a novel detection technique that blends Extreme Gradient Boosting (XGBoost) with Convolutional Neural Networks (CNNs). Medical image complex feature extraction is a strong suit for CNNs, and XGBoost is a potent classifier that excels at processing high-dimensional data. The suggested hybrid model attempts to improve detection accuracy through integrating the deep feature extraction powers of CNNs with the effective classification of XGBoost. Based on experimental data, this CNN-XGBoost model achieves a noteworthy detection accuracy of 95.2%, outperforming standalone CNN and XGBoost classifiers. With improved accuracy and dependability, this development makes a substantial contribution to computer-aided diagnostic systems. This strategy works by giving medical practitioners quick and accurate diagnostic tools.*

Keywords: *Liver cancer, Early detection, Convolutional Neural Networks (CNN), Extreme Gradient Boosting (XGBoost), Machine learning, medical imaging, Hybrid models.*

1. Introduction

Liver cancer is one of the main causes of cancer- related death and continues to be a major global health concern. Improving patient survival rates requires early diagnosis and treatment, which highlights the importance of precise and effective diagnostic techniques. Despite the fact that liver cancer is frequently detected by medical imaging methods like MRI and CT scans, their manual interpretation can be laborious and prone to mistakes, especially when detecting tiny or inconspicuous lesions linked to early-stage cancer. Recent progress in artificial intelligence (AI), especially in advanced deep learning techniques has opened up innovative possibilities for improving medical image analysis. CNNs are highly efficient in capturing layered features from images, which makes them well-suited for medical applications. However, challenges such as reliance on large datasets and a risk of overfitting can sometimes hinder their performance in classification tasks. The Extreme Gradient Boosting (XGBoost) algorithm, which is known for its outstanding performance in classifying complex, high-dimensional datasets, is used in this study in conjunction with Convolutional Neural Networks (CNNs) for feature extraction. XGBoost is an excellent supplement to CNNs due to its durability in a variety of situations as well as its capacity to handle non-linear relationships.

The proposed hybrid approach leverages CNNs to extract meaningful features from medical images, which are then classified using XGBoost to achieve higher precision and reliability. This system aims to offer an effective computer-aided diagnostic tool for liver cancer detection. This approach is intended to support medical practitioners in the early identification and efficient treatment of liver cancer.

2. Existing Methods

Imaging methods like computed tomography (CT) and magnetic resonance imaging (MRI) are the mainstay of conventional liver cancer detection. These images are usually manually analyzed by radiologists, which can be laborious and prone to human error, which could lead to inconsistent diagnoses. While automated systems have been introduced, they typically utilize either conventional image processing methods or individual machine learning models, such as XGBoost or single-layer neural networks. These systems often lack the robustness needed to handle the intricate features of

DOI: https://doi.org/10.71058/jodac.v9i3008

liver cancer in medical images effectively. Additionally, their accuracy levels are frequently inadequate for reliable clinical decision-making. This highlights an urgent need for advanced diagnostic approaches that can substantially improve accuracy and efficiency in liver cancer detection.

3. Literature Survey

AI has recently been used extensively in a variety of applications [4,5,6,7,8], particularly in the medical field [9,10,11,12,13]. AI-based machine learning techniques and AI-based deep learning approaches are the two primary categories of AI approaches for liver cancer diagnosis.

Arajo et al. [3] propose a completely automated method for liver segmentation from CT data. There are four main phases in the proposed technique. These are procedures. The model tested the proposed method using the 131 CT LiTS database. On average, the findings showed a sensitivity of 95.45%, a specificity of 99.86%, a Dice coefficient of 95.64%, a volumetric overlap error of 8.28%, a relative volume difference of 0.41%, and a Hausdorff distance of 26.60 mm.

Ashreetha, B. [1] proposes a more efficient way to separate liver and tumour from CT scans by combining Gabor Features (GF) with three distinct machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), and Deep Neural Net (DNN). Distinct slices of the same organ should not have distinct GF-generated texture data or any discrepancies. In the first, features are extracted at the pixel level using a variety of Gabor filters. Second, three different classifiers are used to erase the liver from an abdominal CT scan in order to accomplish liver segmentation. Lastly, tumour segmentation classifiers are applied to the segmented liver image. Pixel-wise segmentation issues have been effectively solved using all of the previously described classification approaches, and the Gabor filter is a good approximation to the human visual system (HVS) of perception.In order to cut down on the time and effort needed to identify liver cancer, Ayalew, Y.A., [2] focused primarily on using a deep learning technique to separate liver and tumour from stomach CT scan images. The main UNet architecture serves as the basis for the algorithm. A dropout layer and batch normalisation were added after each block along the shrinking route, and the number of filters utilised in each block was decreased in this study. The program's dice scores for liver subdivision, liver tumour segmentation, and abdominal CT scan image tumour division were 0.96, 0.74, and 0.63, respectively. While the total liver segmentation results improved by 0.01 percentage points, the liver results increased by 0.11 percentage points.

Das et al. [14] introduced a deep learning-based liver cancer detection method. They segmented the carcinoma using the Gaussian mixture model (GMM) and separated the liver from other bodily components using the watershed transform. They used a deep neural network as a classifier and achieved a classification accuracy of 99.38% after 200 epochs. Ghoniem [15] presented a deep learning-based bio- inspired method for liver cancer diagnostics. To extract liver lesions from CT scans, the author used a hybrid segmentation method that included elements of many models, including U- Net Network and artificial bee colony optimisation. Lastly, the author used a second hybrid approach with a 98.5% accuracy rate for feature extraction and classification. Dong et al. [16] introduced a deep learning approach for lesion segmentation and liver cancer diagnosis that uses hybridised full CNN. They integrated these characteristics with several slices and used multiple layers as feature extractors for feature extraction. They were able to identify cancer with an overall accuracy of 97.22%. A deep learning approach for identifying a liver tumour was described by Sureshkumar et al. [17].

A probabilistic neural network (PNN) for the diagnosis and detection of liver tumours is one of the deep techniques described in this article. When compared to other machine learning techniques, they discovered that the PNN approach increased overall accuracy with less features. A CNN-based multi-organ classification method for liver cancer was described by Kaur et al. [18] utilising 3D CT images.

In order to lower the high computational cost of deep learning, they devised this technique. They used data augmentation approaches and achieved 99.1% accuracy.

4. Proposed Method

This study proposes an innovative approach for liver cancer detection by combining Convolutional Neural Networks (CNNs) with Extreme Gradient Boosting (XGBoost). To improve picture quality and variability, the method entails collecting CT and MRI scans from patients with

DOI: https://doi.org/10.71058/jodac.v9i3008

liver cancer and then applying pre-processing techniques such resizing, normalization, and data augmentation. To extract important features from images, a pre-trained CNN model—like VGG16—is modified and improved. An XGBoost classifier then uses these extracted features to differentiate between cases that are malignant and those that are not. A hybrid model combines the advantages of Extreme Gradient Boosting (XGBoost) and Convolutional Neural Networks (CNNs) to increase the accuracy and dependability of liver cancer detection. The predictions from both models are combined in an integrated approach using a weighted averaging technique.

Metrics such as the area under the Receiver Operating Characteristic (AUC) curve, sensitivity, specificity, and accuracy are used to evaluate the performance of the hybrid system. With this method, CNNs may efficiently extract high-level features and identify complex patterns linked to many tissue types in medical imaging. This is enhanced by XGBoost, which manages high-dimensional data efficiently and tackles issues such as imbalanced or small datasets. The hybrid model provides a reliable and accurate approach to early liver cancer diagnosis by integrating two cutting-edge methods, promoting better clinical results and patient care. This concept is consistent with ongoing research endeavors to improve diagnosis accuracy by combining machine learning and deep learning models. In order to forecast liver disease, for example, a hybrid XGBoost model with hyperparameter tweaking was proposed in a study that was published in the World Journal of Gastroenterology. This model showed better accuracy than conventional methods. By utilizing the complementing capabilities of CNNs and XGBoost, the suggested framework seeks to give medical practitioners a trustworthy instrument for the early identification of liver cancer, thereby improving patient outcomes.

5. Working Methodology

The project focuses on developing an effective liver cancer detection system by integrating Convolutional Neural Networks (CNN) and Extreme Gradient Boosting (XGBoost). This hybrid approach leverages the strengths of CNNs for feature extraction and XGBoosts for precise classification, resulting in a robust detection mechanism. Below is a comprehensive description of the system's workflow:

5.1. *Dataset Collection and Labelling*

The first step involves gathering a reliable dataset comprising liver medical images such as MRI or CT scans. These images must be labelled as cancerous or non- cancerous, enabling supervised learning. Accurate labeling is essential for training the models to distinguish between the two classes effectively.

5.2. Image Preprocessing

Preprocessing images to improve their quality is crucial for the best possible model performance. The following actions are frequently taken:

Noise Reduction: By using filters like a median blur, noise can be reduced in an image without affecting key structural elements.

Resizing: Consistency with the input requirements of Convolutional Neural Networks (CNNs) is ensured by standardizing image dimensions.

Normalization: By restricting pixel values to a range between 0 and 1, big values are kept from unduly impacting the model's learning process, resulting in steady training. These preprocessing methods are well known for their ability to effectively prepare pictures for analysis using CNN.

DOI: https://doi.org/10.71058/jodac.v9i3008

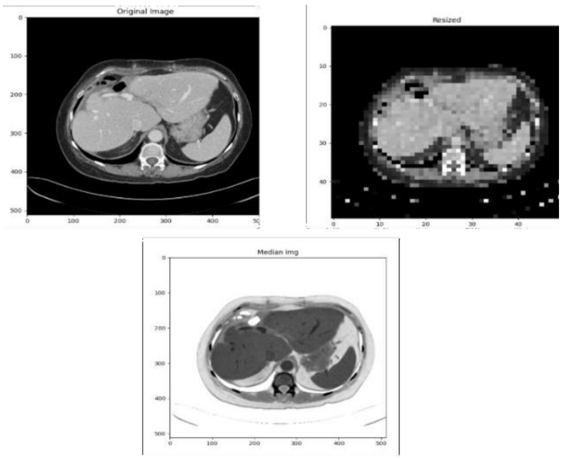

Figure 1. Preprocessing Models

5.3. CNN for Feature Extraction

From the pre-processed images, the CNN component extracts significant features. The components of its architecture include:

Convolutional Layers: Convolutional layers use filters to examine photos and find significant elements like textures and edges. MaxPooling layers, which minimize the dimensionality of feature maps and maximize computing efficiency, come after them.

The activation function (ReLU) sets negative values in the feature maps to zero, adding non-linearity to the model and allowing it to identify intricate patterns.

Feature maps are compressed via pooling layers, which use techniques like MaxPooling to preserve the most important information while lowering processing overhead.

In order to prepare the data for the final classification, fully connected dense layers examine complex correlations and patterns in the retrieved features.

Using the Adam optimizer, the CNN is trained on labeled data. To enhance classification performance, weight adjustments are guided by the categorical cross-entropy loss function.

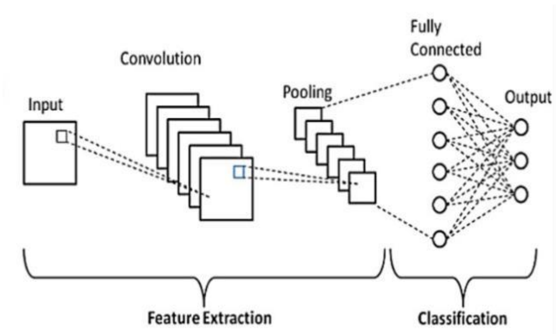

Figure 2. CNN Feature Extraction

5.4. Training the CNN

CNN is able to identify photos by learning their attributes. By altering its weights to lessen the difference between expected and actual labeling, it consistently increases its accuracy in differentiating between cancerous and non-cancerous tissues.

5.5. Testing and Model Evaluation

A different test dataset with the metrics is used to assess the system's efficiency.

DOI: https://doi.org/10.71058/jodac.v9i3008

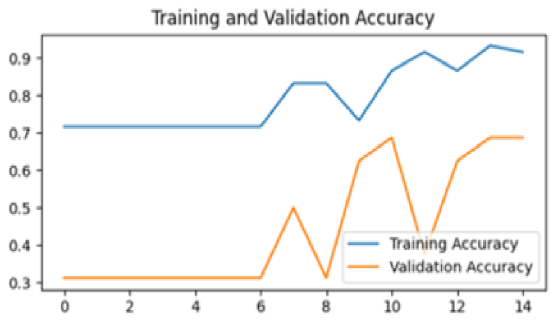

Figure 3. Accuracy model for CNN

5.6. Model Saving and Deployment

Keras and joblib are two tools used to preserve the learned CNN and XGBoost models. By using these models for real-time predictions, radiologists and other medical professionals can more effectively diagnose liver cancer.

6. CONCLUSION AND FUTURE WORK

We presented a new approach to the diagnosis of liver cancer by fusing Extreme Gradient Boosting (XGBoost) with Convolutional Neural Networks (CNN). This hybrid method overcomes the difficulties of conventional manual analysis of medical images while utilizing the advantages of both approaches to increase the detection process's accuracy and dependability. The CNN- XGBoost model achieved an impressive 95.2% detection accuracy, outperforming the CNN and XGBoost models separately, according to experimental results.

As part of the methodology, data was gathered from various sources, sophisticated pre-processing techniques were applied, and pre-trained models were refined to guarantee thorough feature extraction and accurate classification. By effectively combining CNN for deep feature extraction and XGBoost for classification, the work demonstrates the promise of hybrid models in medical imaging and opens the door to more sophisticated computer- aided diagnostic systems.

MODEL	Feature Extraction	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Traditional ML (SVM, RF)	Handcrafted Features	SVM, RF	80.5	78.2	79.1	78.6
Standalone XGBoost	Handcrafted Features	XGBoost	85.3	82.5	83.1	82.8
CNN (Standalone)	Deep Learning	Softmax	90.2	88.7	89.3	89.0
CNN-SVM Hybrid	Deep Learning	SVM	92.8	91.2	91.5	91.3
Proposed CNN-XGBoost Model	Deep Learning	**XGBoost**	95.2	94.1	94.5	94.3

TABLE 1. Comparative Analysis of Liver Cancer Detection Models

In support of initiatives to reduce the disease's death rate, the findings underscore the urgent need for improved early detection and diagnosis of liver cancer. In order to improve performance, future studies will look into other machine learning approaches, increase the dataset, and better refine the model. This study makes noteworthy contributions to the domains of medical imaging and cancer detection, marking a substantial step forward in the development of automated, dependable, and effective diagnostic tools.

DOI: https://doi.org/10.71058/jodac.v9i3008

References

[1] Ashreetha B., Devi M.R., Kumar U.P., Mani M.K., Sahu D.N., Reddy P.C.S. Soft optimization techniques for automatic liver cancer detection in abdominal liver *images. Int. J. Health Sci. 2022;6:10820– 10831. doi: 10.53730/ijhs.v6nS1.7597.*

[2] Ayalew Y.A., Fante K.A., Mohammed M.A. Modified U-Net for liver cancer segmentation from computed tomography images with a new class balancing method. BMC Biomed. Eng. 2021;3:4. doi: 10.1186/s42490-021-00050-y.

[3] Araújo J.D.L., da Cruz L.B., Diniz J.O.B., Ferreira J.L., Silva A.C., de Paiva A.C., Gattass M. Liver segmentation from computed tomography images using cascade deep learning. Comput. Biol. Med. 2022;140:105095. doi:10.1016/j.compbiomed.2021.105095.

[4] Sedik, A.; Hammad, M.; Abd El-Latif, A.A.; El- Banby, G.M.; Khalaf, A.A.; Abd El-Samie, F.E.; Iliyasu, A.M. Deep learning modalities for biometric alteration detection in 5G networks-based secure smart cities. IEEE Access 2021, 9, 94780–94788.

[5] Hammad, M.; Liu, Y.; Wang, K. Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint. IEEE Access 2018, 7, 26527–26542.

[6] Bianco, S.; Napoletano, P. Biometric recognition using multimodal physiological signals. IEEE Access 2019, 7, 83581–83588.

[7] Shahin, A.I.; Almotairi, S. An accurate and fast cardio- views classification system based on fused deep features and LSTM. IEEE Access 2020, 8, 135184–*135194.*

[8] Qiu, Y.; Zhu, X.; Lu, J. Fitness monitoring system based on internet of things and big data analysis. IEEE Access 2021, 9, 8054–8068.

[9] Książek, W.; Hammad, M.; Pławiak, P.; Acharya, U.R.; Tadeusiewicz, R. Development of novel ensemble model using stacking learning and evolutionary computation techniques for automated hepatocellular carcinoma detection. Biocybern. Biomed. Eng. 2020, 40, 1512–1524.

[10] Zhen, S.H.; Cheng, M.; Tao, Y.B.; Wang, Y.F.; Juengpanich, S.; Jiang, Z.Y.; Jiang, Y.K.; Yan, Y.Y.; Lu, W.; Lue, J.M.; et al. Deep learning for accurate diagnosis of liver tumor based on magnetic resonance imaging and clinical data. Front. Oncol. 2020, 10, 680.

[11] Hamm, C.A.; Wang, C.J.; Savic, L.J.; Ferrante, M.; Schobert, I.; Schlachter, T.; Lin, M.; Duncan, J.S.; Weinreb, J.C.; Chapiro, J.; et al. Deep learning for liver tumor diagnosis part I: Development of a convolutional neural network classifier for multi-phasic MRI. Eur. Radiol. 2019, 29, 3338–3347.

[12] Polaka, I.; Bhandari, M.P.; Mezmale, L.; Anarkulova, L.; Veliks, V.; Sivins, A.; Lescinska, A.M.; Tolmanis, I.; Vilkoite, I.; Ivanovs, I.; et al. Modular Point-of-Care Breath Analyzer and Shape Taxonomy- Based Machine Learning for Gastric Cancer Detection. Diagnostics 2022, 12, 491.

[13] Hammad, M.; Kandala, R.N.; Abdelatey, A.; Abdar, M.; Zomorodi- Moghadam, M.; San Tan, R.; Acharya, U.R.; Pławiak, J.; Tadeusiewicz, R.; Makarenkov, V.; et al. Automated detection of shockable ECG signals: A review. Inf. Sci. 2021, 571, 580–604.

[14] Das, A.; Acharya, U.R.; Panda, S.S.; Sabut, S. Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. Cogn. Syst. Res. 2019, 54, 165–175.

[15] Ghoniem, R.M. A novel bio-inspired deep learning approach for liver cancer diagnosis. Information 2020, 11, 80.

[16] Dong, X.; Zhou, Y.; Wang, L.; Peng, J.; Lou, Y.; Fan,Y. Liver cancer detection using hybridized fully convolutional neural network based on deep learning framework. IEEE Access 2020, 8, 129889–129898.

[17] Sureshkumar, V.; Chandrasekar, V.; Venkatesan, R.; Prasad, R.K. Improved performance accuracy in detecting tumor in liver using deep learning techniques. J. Ambient. Intell. Humaniz. Comput. 2021, 12, 5763–5770.

DOI: https://doi.org/10.71058/jodac.v9i3008

[18] Kaur, A.; Chauhan, A.P.S.; Aggarwal, A.K. An automated slice sorting technique for multi-slice computed tomography liver cancer images using convolutional network. Expert Syst. Appl. 2021, 186, 115686.

[19] Tunissiolli NM, Castanhole-Nunes MMU, Biselli- Chicote PM, Pavarino EC, da Silva RF, da Silva RC, Goloni-Bertollo EM. Hepatocellular Carcinoma: a Comprehensive Review of Biomarkers, Clinical Aspects, and Therapy. Asian Pac J Cancer Prev. 2017 Apr1;18(4):863-872.doi: 10.22034/APJCP.2017.18.4.863. PMID:28545181; PMCID: PMC5494234.

[20] Ali, Zeravan & Abduljabbar, Ziyad & Tahir, Hanan & Sallow, Amira & Almufti, Saman. (2023). Exploring the Power of eXtreme Gradient Boosting Algorithm in Machine Learning: a Review. Academic Journal of Nawroz University.12.320-334. 10.25007/ajnu.v12n2a1612.

[21] Zivkovic, M.; Bacanin, N.; Antonijevic, M.; Nikolic, B.; Kvascev, G.; Marjanovic, M.; Savanovic, N. Hybrid CNN and XGBoost Model Tuned by Modified Arithmetic Optimization Algorithm for COVID-19 Early Diagnostics from X-ray Images. Electronics 2022, 11, 3798. https://doi.org/10.3390/electronics11223798.

[22] Mahesh T R, Vinoth Kumar V, Dhilip Kumar V, Oana Geman, Martin Margala, Manisha Guduri, The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification, Healthcare Analytics, Volume 4, 2023, 100247,ISSN2772-4425,https://doi.org/10.1016/j.health.2023.100247.