



Article

Deep Learning for Automated Detection of Periportal Fibrosis in Ultrasound Imaging: Improving Diagnostic Accuracy in *Schistosoma mansoni* Infection

Alex Mutebe ^{1,2,*} , Bakhtiyar Ahmed ³ , Agnes Natukunda ¹, Emily Webb ^{1,4}, Andrew Abaasa ¹, Simon Mpooya ⁵, Moses Egesa ^{1,6,7}, Ayoub Kakande ¹, Alison M. Elliott ¹ and Samuel O. Danso ⁸

- ¹ Medical Research Council/Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine, Uganda Research Unit, Entebbe P.O. Box 49, Uganda
² Department of Computing, University of Essex, Colchester CO4 3SQ, UK
³ Department of Computer Science, Kingston University, London KT1 2EE, UK
⁴ Department of Infectious Disease Epidemiology and International Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
⁵ Division of Vector Borne and Neglected Tropical Diseases, Ministry of Health, Kampala P.O. Box 7272, Uganda
⁶ Department of Infection Biology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
⁷ Uganda Virus Research Institute, Entebbe P.O. Box 49, Uganda
⁸ School of Computer Science and Engineering, University of Sunderland, St Peters Campus, Sunderland SR6 0DD, UK; sam.danso@sunderland.ac.uk
* Correspondence: alex.mutebe@lshtm.ac.uk

Abstract

This study investigates advanced deep learning methods to improve the detection of periportal fibrosis (PPF) in medical imaging. *Schistosoma mansoni* infection affects over 54 million individuals globally, predominantly in sub-Saharan Africa, with around 20 million experiencing chronic complications. PPF, present in up to 42% of these cases, is a leading outcome of chronic liver disease, significantly contributing to morbidity and mortality. Early and accurate detection is critical for timely intervention, yet conventional ultrasound diagnosis remains highly operator-dependent. We adapted and trained a convolutional neural network (CNN) using ultrasound images to automatically identify and classify PPF severity. The proposed approach achieved a diagnostic accuracy of 80%. Sensitivity and specificity reached 84% and 76%, respectively, demonstrating robust generalisability across varying image qualities and acquisition settings. These findings highlight the potential of deep learning to reduce diagnostic subjectivity and support scalable screening programmes. Future work will focus on validation with larger datasets and multi-class fibrosis grading to enhance clinical utility.

Keywords: chronic liver disease; convolutional neural networks; deep learning; diagnostic accuracy; medical imaging; periportal fibrosis; *Schistosoma mansoni*; ultrasound



Academic Editor: Pedro Couto

Received: 4 November 2025

Revised: 27 November 2025

Accepted: 3 December 2025

Published: 21 December 2025

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Intestinal schistosomiasis, caused by the blood fluke *Schistosoma mansoni*, is a major public health concern, affecting approximately 54 million people annually, primarily in sub-Saharan Africa [1]. The infection leads to intestinal schistosomiasis, with pathological manifestations arising from the formation of granulomas around eggs that become lodged in the liver. Granuloma formation leads to PPF, a severe complication affecting a significant proportion of infected individuals, particularly in sub-Saharan Africa [2]. In Uganda, S.

mansoni infection affects up to 70% of the population in endemic regions, with a particularly high prevalence of PPF observed in communities along the shores of Lake Albert and Lake Victoria [3]. PPF is a common manifestation of chronic liver diseases and significantly impacts morbidity and mortality [4]. Early detection of periportal fibrosis is crucial for timely intervention and the potential for reversibility.

Currently, non-invasive diagnostic imaging methods such as ultrasound, CT, MRI, and elastography are used to assess and detect liver damage due to chronic schistosomiasis [5]. However, these methods have limitations. Conventional scoring systems for liver fibrosis based on these imaging techniques are often time-consuming, subjective, and semi-quantitative, leading to variability in interpretation and potential diagnostic inaccuracies [6].

As illustrated in Figure 1, ultrasound imaging can visualize characteristic features of PPF. However, interpretation still relies heavily on the sonographer's expertise, which introduces variability and limits scalability in low-resource settings. These limitations stem from the inherent subjectivity in interpreting imaging features and the reduced sensitivity of these methods in identifying the subtle signs of early-stage fibrosis.

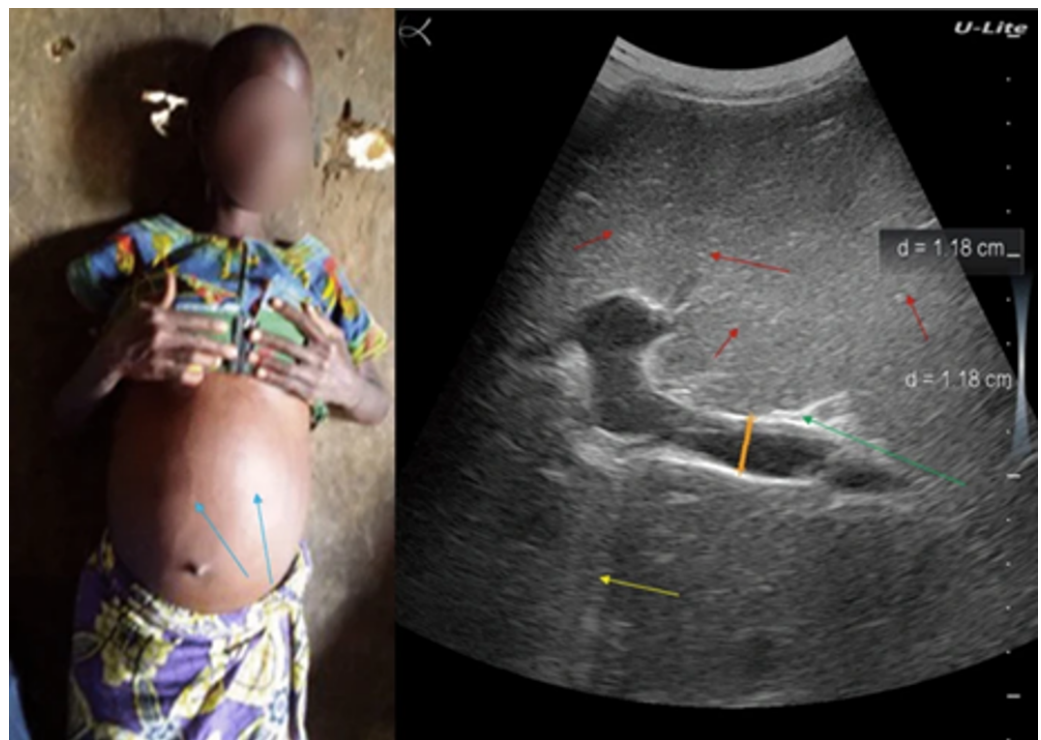


Figure 1. A clinical photograph showing abdominal distension in a patient with PPF (left), and the corresponding ultrasound image (right) with heterogeneous liver architecture, decreased portal vein wall definition with a positive PPF diagnosis [7].

Machine learning (ML), a branch of artificial intelligence (AI), enables computers to recognise patterns in data, supporting prediction and decision-making [8]. Positioned within the broader AI framework, ML plays a pivotal role in the development of intelligent diagnostic systems in healthcare. A computer system trained on thousands of medical images can learn to recognise subtle pathological patterns linked to liver disease and, without explicit programming, continuously improves its diagnostic performance through exposure to data and experience.

Traditional machine learning techniques have been instrumental in advancing medical imaging, supporting tasks such as classification, segmentation, and registration [9]. These methods rely on manual feature extraction, where domain experts identify characteristics such as texture, shape, or intensity, and feed them into algorithms like Support Vector

Machines (SVMs) for classification [10,11]. However, this process is inherently subjective and sensitive to image variations, often leading to inconsistent diagnostic outcomes [12]. Traditional approaches also struggle with the high dimensionality and subtle patterns present in medical images, limiting their ability to detect complex pathological features [13]. Even established models like SVMs have shown inefficiencies in these contexts compared to newer, data-driven deep learning approaches [14].

Traditional machine learning techniques have made significant contributions to medical image analysis; however, their limitations have led to the development of more advanced methods, notably deep learning. Deep Learning (DL) is a subset of machine learning that uses deep neural networks (DNNs) with multi-layered architectures designed to automatically learn complex patterns and representations from large amounts of data. Inspired by the human brain (Figure 2), DL algorithms overcome limitations and reveal new possibilities in medical imaging [15].

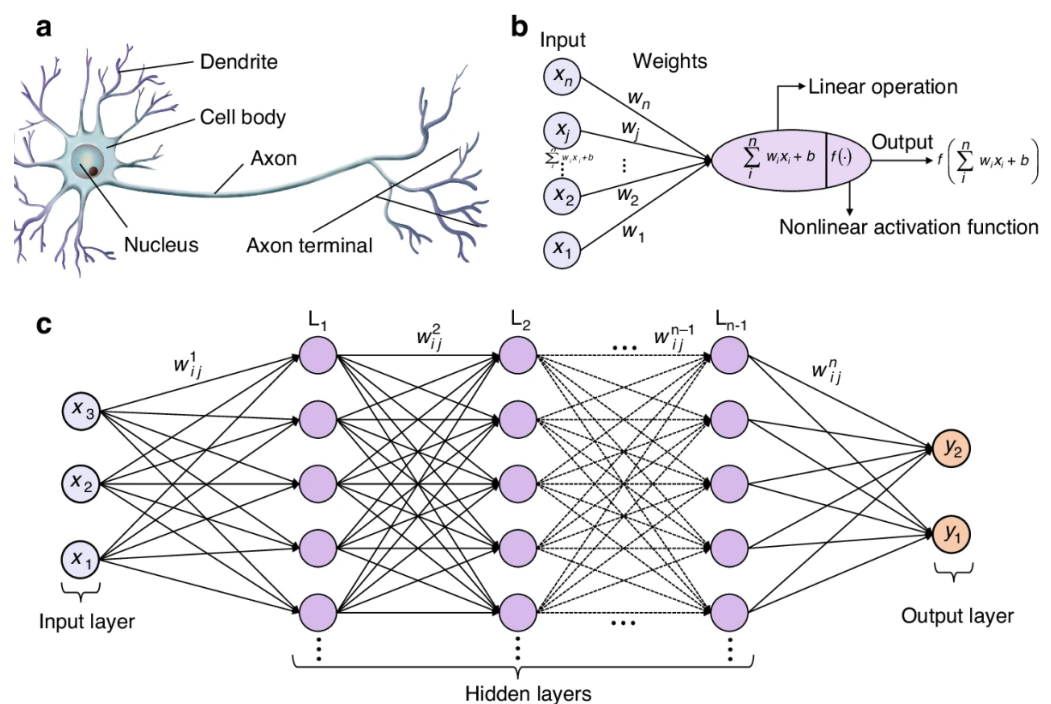


Figure 2. (a) A natural neuron forms the basis of biological neural processing. (b) An artificial neuron models this with a weighted sum of inputs and a nonlinear activation. (c) A deep neural network consists of multiple layers of such neurons, enabling automatic learning of complex patterns from image data [16].

DNNs automate feature extraction and handle the complexity of medical image data, unlike traditional methods dependent on manual feature engineering (Figure 3). DNN algorithms autonomously learn complex patterns and representations from raw image data, identifying subtle details and relationships that traditional techniques might miss [12].

Deep Neural Networks excel due to their ability to detect complex visual patterns, generate consistent and objective measurements, and eliminate the need for manual feature selection. These attributes have proven instrumental in enhancing diagnostic accuracy for PPF detection [15].

Whereas DL has been applied to liver fibrosis detection more broadly [17], few studies have specifically addressed schistosomiasis-related PPF using the Niamey protocol, a standardised set of ultrasound guidelines for the assessment of schistosomiasis-related morbidity, particularly PPF caused by *Schistosoma mansoni* [18,19]. This study aims to fill this gap by implementing and evaluating a DL-based approach for automated detection of

PPF in ultrasound images, providing an objective and scalable tool for early diagnosis in endemic regions.

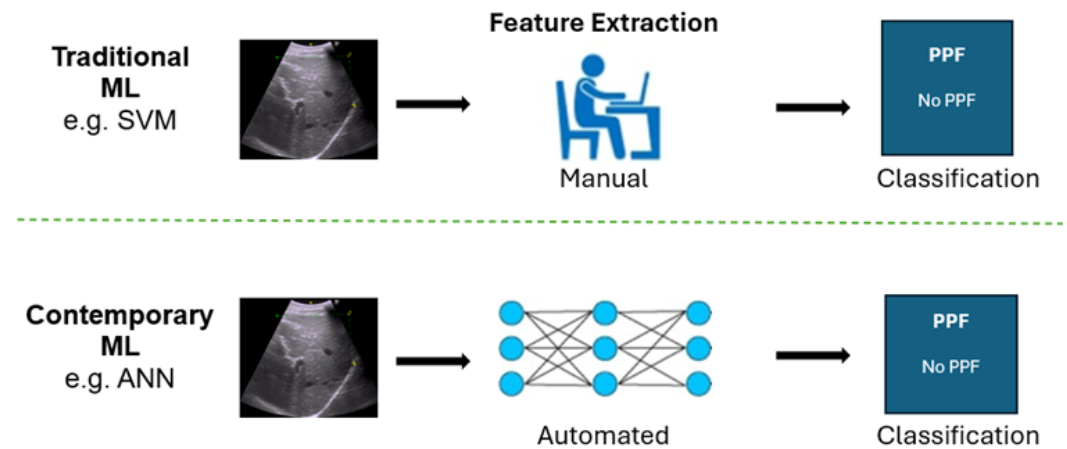


Figure 3. Comparison of the contrast in feature selection approaches between traditional methods and DNNs.

2. Materials and Methods

2.1. Dataset Source

This study utilised a comprehensive liver ultrasound image dataset from a case-control study conducted by the Uganda Schistosomiasis Multidisciplinary Research Centre (U-SMRC) [20]. Data were collected between October 2023 and June 2024 from adult participants living in communities near Lake Victoria and Lake Albert, two distinct epidemiological settings [21]. The study aimed to investigate risk factors associated with severe schistosomal morbidity by comparing individuals with advanced disease (cases) to those without or with mild infection (controls). Liver ultrasound images were part of the inclusion criteria to assess schistosomiasis-related PPF.

The liver ultrasound images were meticulously annotated by an experienced study sonographer using the Niamey protocol, a standardised ultrasonography protocol, which is fundamental for assessing schistosomiasis related morbidity, particularly hepatic morbidity caused by *Schistosoma mansoni* [18,19].

The Image Pattern Score (IPS), derived from the Niamey protocol, provided a standardised approach for evaluating liver damage resulting from schistosomiasis. The score incorporates key sonographic indicators including liver surface nodularity, periportal thickening, parenchymal echogenicity, signal attenuation, and intrahepatic nodules or masses to assess the severity of PPF [22]. Trained sonographers systematically evaluated these features and assigned an IPS.

In this study, we employed image classification as the primary task, framed within the paradigm of supervised learning. In supervised learning, models are trained on labelled datasets, where each input image is paired with a known outcome, enabling the system to learn mappings from input features to target labels [23]. Ultrasound images were labelled according to the IPS assigned by the study's radiologist following the Niamey protocol. Specifically, we designated images with IPS of 2 or higher as cases and those with IPS of 0 or 1 as controls. The CNN was trained to identify subtle ultrasound features associated with PPF, conditioned on these labels, and subsequently applied to unseen scans for classification.

The dataset source comprised 791 liver ultrasound images (197 cases and 594 controls) obtained from adults aged 18 to 50 years. However, reliance on secondary data from a resource-constrained setting, together with issues such as untraceable images arising

from inconsistencies in the image identification numbers stored on the ultrasound device, resulted in inconsistent data quality and restricted the number of verifiable samples. Furthermore, the dataset exhibited a substantial class imbalance, with substantially more non-diseased than diseased cases. Under these constraints, only 200 images (100 non-disease cases and 100 disease cases) could be reliably verified and included to form a balanced dataset for binary classification. Consequently, the sample size was determined by data accessibility and verification feasibility rather than statistical estimation. Of the 200 labelled images, 80% were used for training and 20% for testing.

All ultrasound images were stripped of identifying information to protect participant privacy, in compliance with the Uganda Data Privacy and Protection Act [24] and the General Data Protection Regulation [25]. Participants in the original USMRC study had given prior consent for their ultrasound data to be used in research, allowing for ethical secondary analysis. The study received approval from the University of Essex ethics committee following the submission of a detailed proposal, including formal consent from USMRC principal investigators.

2.2. Data Extraction and Pre-Processing

As illustrated in Figure 4, 3D Slicer was used to extract and anonymise DICOM ultrasound frames of varying sizes, reflecting differences in body habitus and imaging parameters. Larger or deeper livers required greater imaging depth and a wider field of view, whereas smaller or shallower livers required less. Variations in probe positioning, transducer settings, gain, and resolution also influenced frame dimensions. To standardize the dataset, all frames were resized to uniform dimensions (32×32 and 128×128 pixels) while preserving the original aspect ratio using padding, ensuring consistency and computational efficiency during model training.

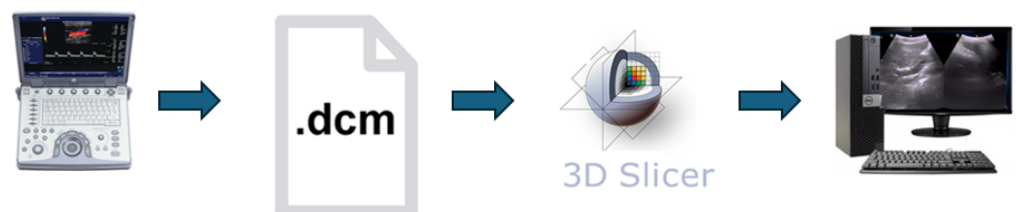


Figure 4. Pipeline illustrating the extraction and storage of ultrasonography (US) images using 3D Slicer 5.6.2 software.

Processed images were exported as PNG files, with all identifiable metadata removed to ensure compliance with privacy regulations and maintain participant anonymity. Several steps were performed to optimize image data for model training. Study IDs were replaced with anonymised labels, and images were categorized with prefixes (e.g., fibrosis_, nofibrosis_) for classification.

Participants were scanned in the supine position after fasting or consuming only water to optimize visualization of abdominal organs. Examinations were conducted using a GE Healthcare Logiq E portable ultrasound system equipped with a 4 MHz curved linear transducer and colour Doppler capability. All images were acquired in B-mode and stored with participant identifiers. Scanning was performed by a Radiological Technologist with over 20 years of experience in diagnosing *Schistosoma mansoni*-related morbidities and applying the Niamey Protocol.

To enhance the robustness and generalisation capability of the CNN, data augmentation techniques were applied during training using the ImageDataGenerator class from Keras. The augmentation strategy introduced random variability into the training images through Random rotations of up to 20° , Horizontal and vertical shifts of up to 20% of the

image width and height, respectively and Random horizontal flips. These transformations simulate common variations in medical or real-world imaging and increase the diversity of the training data. As a result, they help to reduce overfitting and improve the model's ability to generalize to unseen data.

Data Normalization was performed to enhance training stability and model performance, all pixel values in the training and test datasets were first cast to 32-bit floating-point format. The mean and standard deviation of the training images were then computed, representing the average brightness and pixel value dispersion, respectively. Normalization was performed using the equation:

$$\text{Normalized pixel} = \frac{\text{Pixel} - \mu}{\sigma + 10^{-7}} \quad (1)$$

where μ and σ denote the mean and standard deviation of the training dataset. A small constant (10^{-7}) was added to the denominator to avoid division by zero. This normalization was also applied to the test dataset using the training set statistics, ensuring consistent preprocessing across datasets and avoiding data leakage.

2.3. CNN Model Implementation

The CNN model in this study was implemented using the Keras Sequential API, a widely adopted high-level deep learning framework built on top of TensorFlow [26]. The implementation was carried out in Python 3.12.3 and utilised several libraries and packages for model construction, training, evaluation, and visualisation. The model layers were constructed using keras.

We used Google Drive, Git 2.34.1, and Google Colab for an efficient model development workflow. Google Colab is a cloud platform which provides access to high-performance hardware such as GPUs and TPUs, significantly accelerating computations compared with typical local machines. This results in faster model training and potentially higher accuracy [27].

The workflow involved several key steps to ensure seamless integration between data storage, version control, and model training. Initially, periportal images were collected on the local computer and stored in a designated Google Drive directory. These images were organized into two folders labelled “fibrosis” and “nofibrosis” for clear categorization. Jupyter notebooks used for data preprocessing, model training, and evaluation were developed and stored in a GitHub repository, enabling version control and collaboration. Finally, these notebooks and the pre-processed data were linked to Google Colab, where the actual model training took place using the pre-processed data and the computational resources provided by the cloud platform.

2.4. VGG16-Inspired CNN Architecture

The choice of a CNN for PPF detection was motivated by its proven effectiveness in image processing tasks. We adopted a model inspired by the VGG16 network, known for its balance between simplicity and performance. This architecture was selected for its ability to extract deep features from ultrasound images while remaining computationally efficient [28–30].

Compared with other models such as ResNet or transformer-based approaches, the VGG16-inspired CNN offers several advantages [31]. Although the dataset used in this study was relatively small, the deep structure of the network enabled effective feature learning while minimizing the risk of overfitting. Its convolutional layers effectively captured detailed and discriminative image features, supporting accurate fibrosis detection. VGG16 and similar architectures have also demonstrated consistent success across a range

of medical imaging applications, including disease classification. Furthermore, the model's computational efficiency made it well-suited to the resource constraints of this study, unlike more complex transformer-based models that demand greater computational power. Overall, the VGG16-inspired CNN provided an optimal balance between performance and efficiency, making it an appropriate choice for the study's objectives.

The CNN used two activation functions at different stages to perform binary PPF image classification. Rectified Linear Unit (ReLU) activation functions were applied to all convolutional and dense layers except the final layer. Defined as $\text{ReLU}(x) = \max(0, x)$, the ReLU function introduces non-linearity by zeroing out all negative input values while preserving positive ones. This facilitates efficient training by mitigating the vanishing gradient problem and enabling the network to learn hierarchical, discriminative features from liver ultrasound images such as texture, edge contrast, and structural anomalies. The output layer consisted of two neurons activated by the hlsigmoid function, defined as:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Each sigmoid unit independently maps its input to a probability range between 0 and 1 (Figure 5), allowing the model to assign class confidence scores for the presence or absence of PPF. Since the classification problem was binary and the dataset had a balanced class distribution, a threshold of 0.5 was applied [32], outputs > 0.5 were interpreted as PPF-positive, and outputs ≤ 0.5 as PPF-negative. In deployment, the higher of the two probabilities was used to determine the predicted class.

Sigmoid Activation Function with Threshold at 0.5

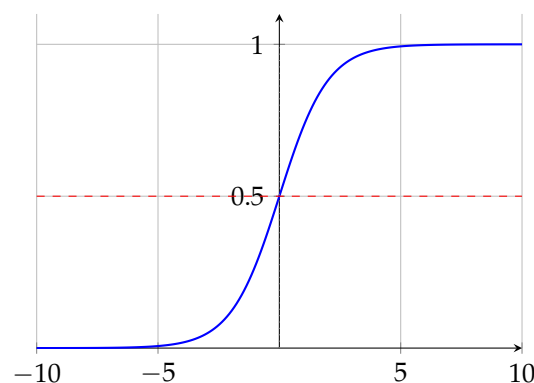


Figure 5. Plot of the sigmoid activation function with a threshold at 0.5.

Together, this activation configuration enabled the model to learn rich representations of image features and translate them into clinically relevant binary predictions.

2.5. Evaluation Metrics

To evaluate model performance, we used standard classification metrics derived from the confusion matrix, including accuracy, precision, recall, F1 score, and specificity. In addition, we assessed the area under the ROC curve (AUC). Together, these metrics provide a comprehensive assessment of both the sensitivity and reliability of the model.

Accuracy represents the proportion of correctly classified instances (both positives and negatives) among the total number of cases. It provides an overall measure of how well the model performs across all classes. It is calculated as:

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

Precision is defined as the proportion of true positive predictions among all instances that were predicted as positive. It is given by:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall, also referred to as sensitivity or the true positive rate, is defined as the proportion of true positive predictions among all actual positive instances. It is calculated using the formula:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

The F1 score is a combined measure of precision and recall, calculated as their harmonic mean. It provides a balance between the two metrics, especially in scenarios where data is imbalanced or when both false positives and false negatives carry significant consequences.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Specificity, also called the true negative rate, measures the proportion of correctly identified negatives among all actual negatives. It is expressed as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

The AUC is a threshold-independent metric that evaluates the model's ability to distinguish between positive and negative classes. It is computed as the area under the ROC curve, where the curve plots the true positive rate against the false positive rate at various thresholds. An AUC of 1.0 indicates perfect classification, whereas 0.5 corresponds to random guessing.

2.6. Model Training

Random seed initialisation was performed using Python's built-in random and numpy modules to ensure reproducibility of training results. Several key hyperparameters were selected and adjusted during the model training process. These adjustments were supported by callbacks that helped fine-tune training based on model performance. Training began by evaluating two input image resolutions (32×32 pixels and 128×128 pixels), cycling through each size to identify the optimal scale for feature extraction. Comparison of the two input resolutions showed that the model trained on 32×32 frames achieved higher accuracy and greater consistency across validation runs than the 128×128 configuration. The network employed the Adam optimizer with an initial learning rate of 1×10^{-3} .

To enhance generalisation and mitigate overfitting, on-the-fly data augmentation was applied using Keras's ImageDataGenerator, which randomly rotated, shifted, and flipped input images during training. Finally, a dropout rate of 0.5 was used in the fully connected layer to reduce overfitting by randomly deactivating half the neurons during each update. In addition to our baseline training setup, we leveraged two key Keras callbacks to enhance model performance. The first, ReduceLROnPlateau, monitors validation loss and automatically reduces the learning rate by a factor of 0.1 after 10 epochs without improvement. This approach helps the optimizer converge more precisely once performance plateaus [33]. The second callback, EarlyStopping, halts training when no further progress is seen over ten epochs. Together, these callbacks streamlined training, prevented overfitting, and secured an optimally tuned model.

In summary, training was performed using the Adam optimiser, selected for its adaptive learning rate and strong empirical performance on complex medical imaging tasks. For activation functions, we applied the ReLU in all hidden layers to introduce non-linearity

while maintaining computational efficiency. The final layer used a sigmoid activation function to output probabilities suitable for binary classification, specifically distinguishing between cases with and without PPF [34]. Although no automated search strategy was applied, the use of well-chosen hyperparameters and responsive callbacks played an important role in shaping the model's learning behaviour and improving overall performance.

3. Results

Table 1 presents the descriptive characteristics of the study population from whom the ultrasound images were collected and used to train and test the CNN model. It summarizes the demographic, anthropometric, and ultrasound IPS for the 200 individuals included in the analysis.

Table 1. Characteristics of study participants stratified by PPF status (N = 200).

Characteristic	No PPF (N = 100)	PPF (N = 100)
Sex		
Female	28 (28%)	18 (18%)
Male	72 (72%)	82 (82%)
Age (years)	29 (25, 37)	30 (24, 39)
Left Liver Lobe Length (cm)	6.90 (6.15, 7.60)	8.15 (7.25, 9.20)
Right Liver Lobe Length (cm)	9.90 (9.40, 10.30)	10.40 (9.90, 11.40)
Inner-to-Inner Diameter of Branch 1 (mm)	2.20 (1.80, 2.50)	2.50 (1.85, 3.10)
Outer-to-Outer Diameter of Branch 1 (mm)	4.30 (3.60, 5.10)	6.50 (5.40, 7.90)
Inner-to-Inner Diameter of Branch 2 (mm)	2.20 (1.80, 2.50)	2.50 (2.20, 3.00)
Outer-to-Outer Diameter of Branch 2 (mm)	4.00 (3.60, 4.85)	6.95 (5.45, 8.10)
Image Pattern Score		
0	88 (88%)	0 (0%)
1	12 (12%)	0 (0%)
2	0 (0%)	54 (54%)
4	0 (0%)	40 (40%)
6	0 (0%)	6 (6.0%)

Values are presented as n (%) for categorical variables and Median (Q1, Q3) for continuous variables.

Initially, a baseline model (Model 1) was developed using a batch size of 16. It achieved a test accuracy of 83% and a test loss of 0.40, showing promising results [35,36].

Figure 6 loss and accuracy curves indicated overfitting, as reflected by stable high validation accuracy but fluctuating validation loss. This suggested that the model memorised training data rather than generalizing well.

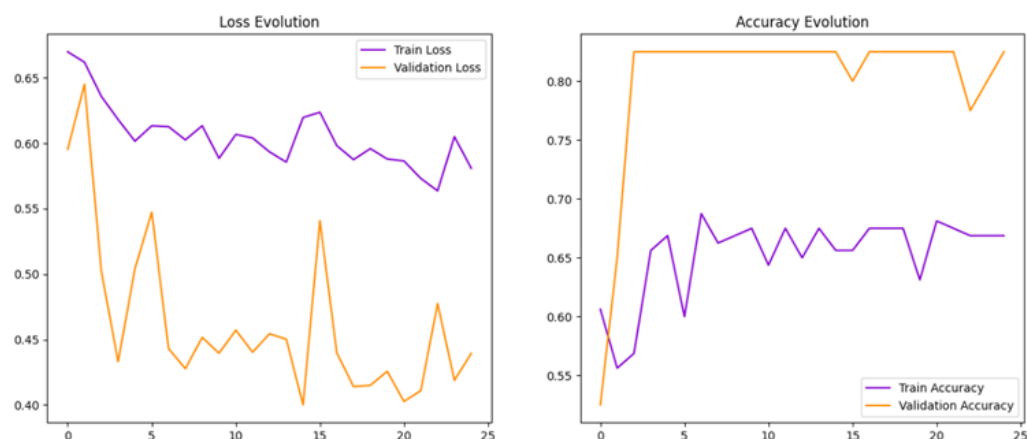


Figure 6. Training and validation curves for Model 1.

To help reduce overfitting, we increased the batch size from 16 to 32 in Model 2. This adjustment aimed to stabilise training and improve the model's ability to generalise.

Model Performance Evaluation

Model 2 achieved a test accuracy of 80% (Table 2), with smoother learning curves observed during training (Figure 7), suggesting improved generalisation. The 95% confidence interval (CI) for accuracy was [77.5%, 97.5%]. The model attained an AUC of 0.87, with a 95% CI of [78%, 99.5%], reflecting strong discriminative ability between classes.

Table 2. Comparison of model performance metrics.

Model	Accuracy	AUC	Precision	Sensitivity	Specificity	F1 Score
Model 1	83	89	73	100	67	84
Model 2	80	87	76	84	76	80

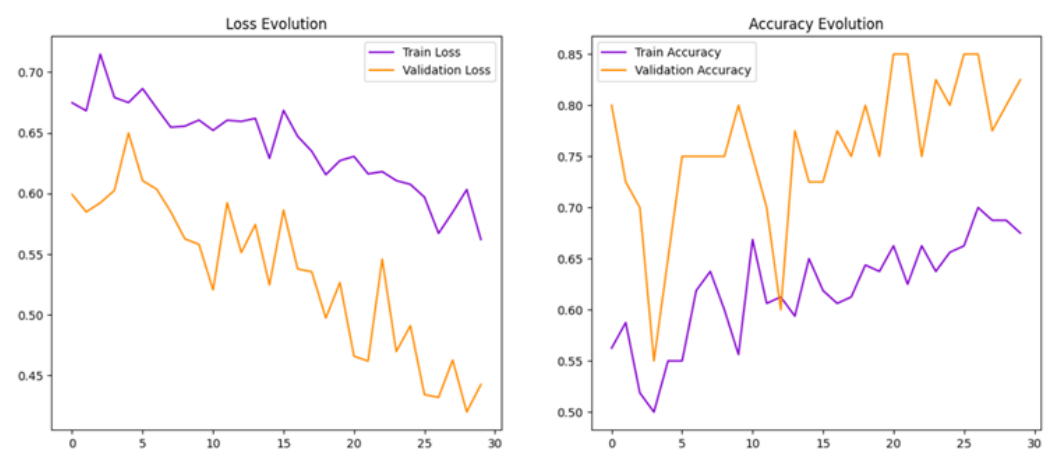


Figure 7. Training and validation curves for Model 2, indicating smoother convergence during training and clearer improvements in generalisation compared to model 1.

Figures 8 and 9 show the confusion matrix for PPF detection, summarizing true positives, false negatives, false positives, and true negatives.

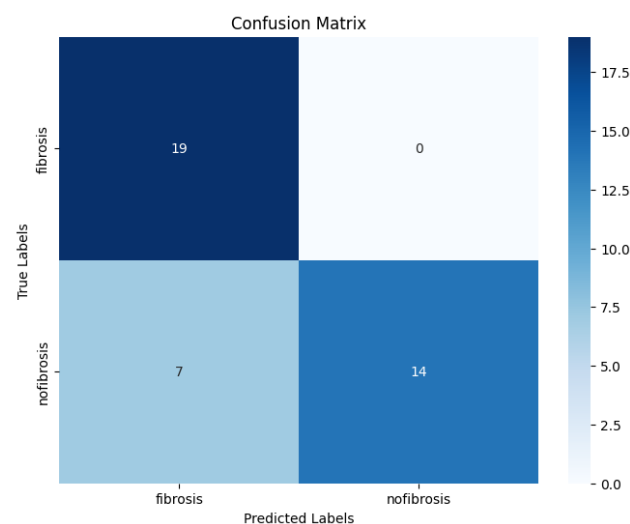


Figure 8. Confusion matrix for Model 1.

Model 2 achieved a precision of 76%, a recall of 84%, and an F1 score of 80%, indicating balanced performance between correctly identifying fibrosis cases and avoiding false positives. Specificity was 76%, showing the model's ability to correctly identify nofibrosis cases.

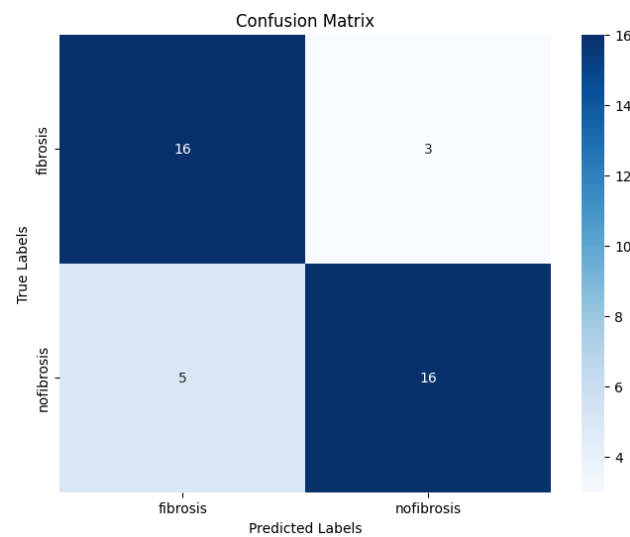


Figure 9. Confusion matrix for Model 2.

The ROC curve for Model 2 is shown in Figure 10, with an AUC of 0.87, demonstrating strong discriminative capability between fibrosis and nofibrosis cases.

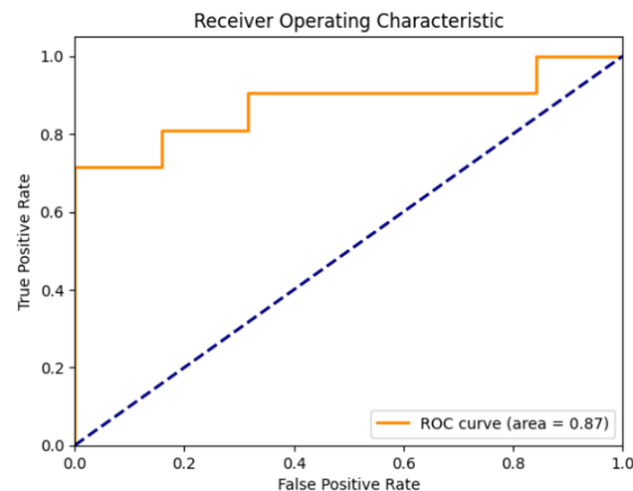


Figure 10. ROC curve for Model 2 with an AUC of 0.87.

In summary, Model 1 achieved higher accuracy, F1 score, and AUC, but showed signs of overfitting, raising concerns about reliability on unseen data. Model 2, though with reduced accuracy and F1 score, demonstrated greater stability across sensitivity and specificity and better resisted overfitting. Overall, Model 1 excelled in headline performance metrics, but Model 2 provided stronger generalisation and consistency, making it more suitable for real world clinical deployment in resource limited settings.

4. Discussion

This study demonstrates that a CNN can effectively detect PPF from ultrasound images. Even at a reduced 32×32 resolution, diagnostic performance remained robust, indicating that essential visual features were preserved. The results highlight the value of a

well-designed preprocessing pipeline that maintains structural integrity while improving computational efficiency. Preserving the original aspect ratio during resizing prevented distortion of anatomical structures, allowing the model to focus on relevant echogenic and textural features. These findings also emphasise the potential of lightweight, resource-efficient models for ultrasound analysis, particularly in field or low-resource settings, and support the feasibility of scalable, automated approaches for PPF screening.

The second model (Model 2), developed through iterative optimisation, achieved strong overall performance, suggesting that deep learning can play a valuable role in improving diagnostic accuracy for liver disease caused by *Schistosoma mansoni*. The results reflect a balanced ability to both identify true positive cases and avoid false positives, a critical consideration in clinical diagnostics. The model's high precision indicates that when fibrosis is predicted, it is usually correct, reducing unnecessary concern or follow-up. At the same time, its ability to capture most positive cases highlights its potential to support earlier identification of patients with fibrosis. Taken together, these findings point to a model that could complement human expertise by providing consistent and reliable diagnostic support.

The area under the curve further reinforces the strength of the classifier, indicating that the model could be confidently applied to distinguish between individuals with and without PPF [37]. One of the main improvements from the baseline model (Model 1) to Model 2 was in addressing overfitting. Initially, while the model achieved strong training performance, the unstable validation loss suggested it had memorised the training data rather than learned general patterns. By adjusting the batch size, implementing early stopping, and applying data augmentation, Model 2 was able to generalise better. The resulting learning curves, though showing mild oscillations, indicated more consistent performance on unseen data. Such fluctuations are common in models trained on limited or noisy datasets, where validation metrics often reflect sampling variability rather than convergence failure [38]. Nevertheless, the overall trajectories demonstrated stable and progressive optimisation. To further mitigate overfitting and improve generalisation, expanding the dataset to capture greater diversity and better represent real-world distributions is recommended [39]. The results reported here, however, reflect the best-performing configuration following iterative adjustment of training parameters.

In clinical practice, diagnosing PPF through ultrasound imaging requires significant expertise and is prone to subjectivity. The use of a CNN model provides an opportunity to standardise this process, reducing reliance on expert interpretation and increasing the consistency of results. This could be especially beneficial in low-resource settings where access to trained radiologists or sonographers is limited.

Moreover, early detection of PPF is crucial, as it can prevent progression to more severe liver complications. An automated tool capable of identifying early signs of fibrosis in routine ultrasound scans could support timely interventions and improve patient outcomes.

The results observed here are comparable to those from similar studies. For instance, Lee et al. (2020) [17] reported strong performance using deep learning on ultrasound data to detect liver fibrosis, similar to the performance achieved in this study. However, this project differs in that it focuses specifically on PPF, rather than general liver fibrosis, and applies the Niamey protocol, which is widely used in field-based diagnostics. This makes the model more applicable to schistosomiasis-endemic settings, particularly in sub-Saharan Africa.

While the results are encouraging, this study has a number of limitations. The dataset included only 200 ultrasound images, which may restrict how well the model generalises to broader populations or different imaging settings. Another limitation is that the initial classification of images was carried out by only one ultrasonographer. Having a second independent reviewer would have reduced the risk of relying on a single person's subjec-

tive perspective. Moreover, no other types of liver disease were included, meaning the procedure may not perform as well if other causes of fibrosis are present.

Future research should therefore aim to train and validate the model on larger and more varied datasets, ideally collected from multiple centres and regions. In addition, future studies should consider stratified or randomised recruitment strategies to minimise bias and improve dataset representativeness. Another direction is extending the model beyond binary classification (fibrosis versus no fibrosis) to predict different levels of fibrosis severity, as defined by the Niamy grading system, which could enhance clinical utility. Future work should incorporate a wider range of hepatic pathologies to ensure the model can distinguish PPF from other causes of liver fibrosis. Exploring alternative architectures, such as ResNet, DenseNet, or EfficientNet, may provide improved accuracy and efficiency compared to VGG16. Ensemble approaches could also be investigated to combine the strengths of multiple models for more robust PPF detection.

Manual hyperparameter tuning was employed in this study to optimize model performance. While this approach allowed careful adjustment based on observed model behavior, we acknowledge that automated methods such as Bayesian optimisation could be explored in future work to further refine the hyperparameters systematically. Future work could also explore the use of interpretability methods such as Gradient-weighted Class Activation Mapping (Grad-CAM), which highlight the regions of an image most influential in the model's decision. This would make the system more transparent and help build clinician confidence in its outputs.

5. Conclusions

This study adds to the growing body of evidence supporting the application of deep learning techniques to routine ultrasound imaging for the detection of PPF. The CNN model developed demonstrated strong diagnostic performance, highlighting its potential to assist clinical decision-making, particularly in schistosomiasis-endemic regions where access to specialised radiological expertise is limited. With further validation, this AI-based approach could help shift fibrosis diagnosis from a subjective process to a more standardised and scalable one, thereby promoting more equitable and timely care for individuals at risk of liver complications due to schistosomiasis.

In summary, our CNN achieved strong diagnostic performance for detecting PPF in ultrasound images. With further validation on larger and more diverse datasets, this approach could support earlier, more consistent diagnosis in endemic regions. We acknowledge that deployment metrics (e.g., inference latency, memory footprint) and comparative evaluation against clinical experts were beyond the scope of this study. Future work will address these aspects, alongside extending beyond binary classification to capture fibrosis severity and improve interpretability for clinical adoption.

Author Contributions: Conceptualization, A.M.; methodology, S.O.D. and B.A.; software, A.M.; validation, S.M.; formal analysis, A.M.; investigation, A.M.; resources, A.A.; data curation, A.K.; writing—original draft preparation, A.M.; writing—review and editing, B.A., A.N., E.W., A.A., A.M.E., S.M., M.E. and A.K.; visualization, A.M.; supervision, S.O.D. and B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval for this secondary analysis of U-SMRC ultrasound data was obtained from the University of Essex Ethics Committee (Ref: GC/127/940; Date: 17 September 2024) following submission of a detailed proposal.

Informed Consent Statement: Informed consent was obtained from all participants in the original U-SMRC study for their ultrasound data to be used in research, allowing for ethical secondary analysis. Written informed consent for publication was not applicable, as individual participants cannot be identified from the data.

Data Availability Statement: The de-identified individual participant data that underlie the results reported in this research are stored in a non-publicly available repository at the MRC/UVRI & LSHTM Uganda Research Unit. Data access is subject to a request and review process. Researchers wishing to access the data should submit a request outlining the data required and the intended use. Requests are reviewed by the Principal Investigator in consultation with the MRC/UVRI & LSHTM data management committee, with oversight from the UVRI and LSHTM ethics committees. In line with the MRC data sharing policy, requests will not be unreasonably refused. Researchers granted access will be required to sign a data sharing agreement restricting use to answering the pre-specified research questions.

Acknowledgments: This work reported in this publication was supported in part by grant number U01AI168609 from the United States National Institute of Allergy and Infectious Diseases.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
PPF	Periportal Fibrosis
US	Ultrasound
IPS	Image Pattern Score
Grad-CAM	Gradient-weighted Class Activation Mapping
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
ReLU	Rectified Linear Unit
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit
SVM	Support Vector Machine
DICOM	Digital Imaging and Communications in Medicine
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
MRC	Medical Research Council
UVRI	Uganda Virus Research Institute
LSHTM	London School of Hygiene & Tropical Medicine
USMRC	Uganda Schistosomiasis Multidisciplinary Research Centre
GDPR	General Data Protection Regulation
PNG	Portable Network Graphics
API	Application Programming Interface

References

1. Anderson, T.J.; Enabulele, E.E. *Schistosoma mansoni*. *Trends Parasitol.* **2021**, *37*, 176–177. [[CrossRef](#)]
2. Gunda, D.W.; Kilonzo, S.B.; Manyiri, P.M.; Peck, R.N.; Mazigo, H.D. Morbidity and mortality due to *Schistosoma mansoni* related periportal fibrosis: Could early diagnosis of varices improve the outcome following available treatment modalities in sub Saharan Africa? A scoping review. *Trop. Med. Infect. Dis.* **2020**, *5*, 20. [[CrossRef](#)]

3. Natukunda, A.; Zirimenya, L.; Nkurunungi, G.; Nassuuna, J.; Nkangi, R.; Mutebe, A.; Corstjens, P.L.; van Dam, G.J.; Elliott, A.M.; Webb, E.L. Pre-vaccination *Schistosoma mansoni* and hookworm infections are associated with altered vaccine immune responses: A longitudinal analysis among adolescents living in helminth-endemic islands of Lake Victoria, Uganda. *Front. Immunol.* **2024**, *15*, 1460183. [CrossRef] [PubMed]
4. Andrianah, G.E.P.; Rakotomena, D.; Rakotondrainibe, A.; Ony, L.H.N.R.N.; Ranoharison, H.D.; Ratsimba, H.R.; Rajaonera, T.; Ahmad, A. Contribution of Ultrasonography in the Diagnosis of Periportal Fibrosis Caused by Schistosomiasis. *J. Med. Ultrasound* **2020**, *28*, 41–43. [CrossRef]
5. Santos, J.C.; Pereira, C.L.D.; Domingues, A.L.C.; Lopes, E.P. Noninvasive diagnosis of periportal fibrosis in *Schistosomiasis mansoni*: A comprehensive review. *World J. Hepatol.* **2022**, *14*, 696. [CrossRef] [PubMed]
6. Masseroli, M.; Caballero, T.; O'Valle, F.; Del Moral, R.M.; Pérez-Milena, A.; Del Moral, R.G. Automatic quantification of liver fibrosis: Design and validation of a new image analysis method: Comparison with semi-quantitative indexes of fibrosis. *J. Hepatol.* **2000**, *32*, 453–464. [CrossRef] [PubMed]
7. Nigo, M.M.; Odermatt, P.; Nigo, D.W.; Salieb-Beugelaar, G.B.; Battegay, M.; Hunziker, P.R. Patients with severe *Schistosomiasis mansoni* in Ituri Province, Democratic Republic of Congo. *Infect. Dis. Poverty* **2021**, *10*, 50–63. [CrossRef]
8. Lee, J.-G.; Jun, S.; Cho, Y.-W.; Lee, H.; Kim, G.B.; Seo, J.B.; Kim, N. Deep learning in medical imaging: General overview. *Korean J. Radiol.* **2017**, *18*, 570–584. [CrossRef]
9. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef]
10. Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [CrossRef]
11. Iwendi, C.; Bashir, A.K.; Peshkar, A.; Sujatha, R.; Chatterjee, J.M.; Pasupuleti, S.; Mishra, R.; Pillai, S.; Jo, O. COVID-19 patient health prediction using boosted random forest algorithm. *Front. Public Health* **2020**, *8*, 357. [CrossRef]
12. Alzubaidi, L. Deep Learning for Medical Imaging Applications. Ph.D. Thesis, Queensland University of Technology, Brisbane City, Australia, 2022.
13. Zhou, S.K.; Greenspan, H.; Davatzikos, C.; Duncan, J.S.; Van Ginneken, B.; Madabhushi, A.; Prince, J.L.; Rueckert, D.; Summers, R.M. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proc. IEEE* **2021**, *109*, 820–838. [CrossRef] [PubMed]
14. Lai, Y. A comparison of traditional machine learning and deep learning in image recognition. *J. Phys. Conf. Ser.* **2019**, *1314*, 012148. [CrossRef]
15. Tang, X. The role of artificial intelligence in medical imaging research. *BJR Open* **2019**, *2*, 20190031. [CrossRef]
16. Fu, T.; Zhang, J.; Sun, R.; Huang, Y.; Xu, W.; Yang, S.; Zhu, Z.; Chen, H. Optical neural networks: Progress and challenges. *Light Sci. Appl.* **2024**, *13*, 263. [CrossRef]
17. Lee, J.H.; Joo, I.; Kang, T.W.; Paik, Y.H.; Sinn, D.H.; Ha, S.Y.; Kim, K.; Choi, C.; Lee, G.; Yi, J.; et al. Deep learning with ultrasonography: Automated classification of liver fibrosis using a deep convolutional neural network. *Eur. Radiol.* **2020**, *30*, 1264–1273. [CrossRef]
18. El Scheich, T.; Holtfreter, M.C.; Ekamp, H.; Singh, D.D.; Mota, R.; Hatz, C.; Richter, J. The WHO ultrasonography protocol for assessing hepatic morbidity due to *Schistosoma mansoni*: Acceptance and evolution over 12 years. *Parasitol. Res.* **2014**, *113*, 3915–3925. [CrossRef]
19. Akpata, R.; Neumayr, A.; Holtfreter, M.C.; Krantz, I.; Singh, D.D.; Mota, R.; Walter, S.; Hatz, C.; Richter, J. The WHO ultrasonography protocol for assessing morbidity due to *Schistosoma haematobium*: Acceptance and evolution over 14 years. Systematic review. *Parasitol. Res.* **2015**, *114*, 1279–1289. [CrossRef]
20. Uganda Schistosomiasis Multidisciplinary Research Center. Building Expertise and Understanding of the Underlying Biological Determinants of Severe Schistosomal Morbidity and Developing Appropriate Interventions for Prevention and Management. Online Resource. 2022. Available online: <https://www.muui.org/usmrc/> (accessed on 18 January 2024).
21. The Compass for SBC. Bilharzia Campaign in Uganda. 2014. Available online: <https://thecompassforsbc.org/sbcc-spotlights/bilharzia-campaign-uganda> (accessed on 5 September 2025).
22. Ockenden, E.S.; Frischer, S.R.; Cheng, H.; Noble, J.A.; Chami, G.F. The role of point-of-care ultrasound in the assessment of schistosomiasis-induced liver fibrosis: A systematic scoping review. *PLoS Neglected Trop. Dis.* **2024**, *18*, e0012033. [CrossRef] [PubMed]
23. IBM Cloud Education. What Is Supervised Learning? Available online: <https://www.ibm.com/think/topics/supervised-learning> (accessed on 18 May 2025).
24. Government of Uganda. Data Protection and Privacy Act. 2019. Available online: <https://ict.go.ug/ims/public/site/documents/Data-Protection-and-Privacy-Act-2019.pdf> (accessed on 20 September 2024).
25. European Union. General Data Protection Regulation (ict). 2016. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504> (accessed on 20 January 2025).

26. Chollet, F. *Deep Learning with Python*; Simon and Schuster: New York, NY, USA, 2021.
27. Kushwaha, U.; Gupta, P.; Airen, S.; Kuliha, M. Analysis of CNN Model with Traditional Approach and Cloud AI based Approach. In Proceedings of the 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 13–15 December 2022; pp. 835–842.
28. Jiang, Z.P.; Liu, Y.Y.; Shao, Z.E.; Huang, K.W. An improved VGG16 model for pneumonia image classification. *Appl. Sci.* **2021**, *11*, 11185. [[CrossRef](#)]
29. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
30. Qassim, H.; Verma, A.; Feinzimer, D. Compressed residual-VGG16 CNN model for big data places image recognition. In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; pp. 169–175.
31. Shaha, M.; Pawar, M. Transfer learning for image classification. In Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 29–31 March 2018; pp. 656–660.
32. Joseph, K. Demystifying Logistic Regression: A Deep Dive. 2024. Available online: <https://medium.com/@josephkiran2001/demystifying-logistic-regression-a-deep-dive-7c41ed510305> (accessed on 18 April 2025).
33. Keras Team. EarlyStopping Callback. 2025. Available online: https://keras.io/api/callbacks/early_stopping/ (accessed on 18 May 2025).
34. Basha, S.H.S.; Dubey, S.R.; Pulabaigari, V.; Mukherjee, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing* **2020**, *378*, 112–119. [[CrossRef](#)]
35. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed.; O'Reilly Media: Sebastopol, CA, USA, 2019.
36. Kufel, J.; Bargieł-Łączek, K.; Kocot, S.; Koźlik, M.; Bartnikowska, W.; Janik, M.; Czogalik, Ł.; Dudek, P.; Magiera, M.; Lis, A.; Paszkiewicz, I. What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. *Diagnostics* **2023**, *13*, 2582. [[CrossRef](#)]
37. Çorbacioğlu, Ş.K.; Aksel, G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turk. J. Emerg. Med.* **2023**, *23*, 195–198. [[CrossRef](#)] [[PubMed](#)]
38. Scikit-Learn Developers. Validation Curves: Plotting Scores to Evaluate Models. Available online: https://scikit-learn.org/stable/modules/learning_curve.html (accessed on 3 October 2025).
39. Ibrahim, M. A Deep Dive Into Learning Curves in Machine Learning. Available online: <https://wandb.ai/mostafaibrahim17/ml-articles/reports/A-Deep-Dive-Into-Learning-Curves-in-Machine-Learning--Vmlldzo0NjA1ODY0> (accessed on 3 October 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.