Liver Disease Detection from CT scan images using Deep Learning and Transfer Learning

MSc Research Project
Data Analytics

Paras Jain
Student ID: 18182119

School of Computing
National College of Ireland

Supervisor: Mr. Hicham Rifai

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Paras Jain		
Student ID:	18182119		
Programme:	Data Analytics	**Year:**	2020
Module:	MSc Research Project		
Supervisor:	Mr. Hicham Rifai		
Submission Due Date:	17/08/2020		
Project Title:	Liver Disease Detection from CT scan images using Deep Learning and Transfer Learning		
Word Count:	8507	**Page Count**	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……

Date: ……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Liver Disease Detection from CT scan images using Deep Learning and Transfer Learning

Paras Jain
18182119

Abstract

The liver is an essential organ in the human body and early diagnosis of various liver diseases can be life-saving. Computer-aided diagnostic systems can assist doctors in the precise diagnosis of liver diseases and eliminate the invasive process of biopsy. Due to complexities present in CT scan images, images are processed using various transformations, augmentation and segmentation techniques. This research project implements SVM, CNN, Inception-v4 and DenseNet-169 models based on deep-learning and transfer-learning with the aim to develop a binary classifier that can accurately differentiate between healthy or normal and unhealthy or abnormal liver. The different techniques of parameter tuning are employed to improve model performance. Resultantly, it is found that CNN and DenseNet-169 models achieve 98.8 % accuracy and 99.66% accuracy, respectively. However, the computational time of DenseNet-16 is significantly higher than that of the CNN model. A lot more work and research remain, but the current results are promising.

1 Introduction

The liver is an essential inner organ of the human body. It performs various metabolic functions such as the production of protein, cholesterol, and bile acid. Its major responsibility is to filter the blood coming from the digestive tract and then transport it to the rest of the body. About 1-1.5 liters of blood is transported to the liver every minute (Lawankar, Sangewar and Gugulothu, 2016). Human survival is not possible without a healthy liver. Excessive consumption of drugs and alcohol is one of the substantial causes of different liver diseases such as cirrhosis, hepatitis virus A, B, C, cancer, etc. and are not related to age. The major noticeable symptoms include weakness, excessive weight loss or gain, and pain in the upper right abdomen. There are broadly two types of liver diseases namely, diffused and focal. When the entire surface of the liver is affected, it is called diffused liver diseases, such as fatty liver and cirrhosis. On the contrary, when a small region of the liver is affected, it is called focal liver diseases, such as cyst, hemangioma, and hepatocellular carcinoma (M. Hassan, Elmogy and Sallam, 2015).

Liver disease can decrease the function of the liver and impact the development of protein, hormones, and nutrients in the human body and the global public health epidemic is frequently increasing. Liver biopsy is considered as a gold standard for the diagnosis of different conditions and fibrosis. Most liver diseases and cancers are caused by alcohol-related cirrhosis and fatty liver disease associated with obesity. Hepatitis B and C viruses are also significant risk factors. The prevalence of liver cancer rose by more than 3% annually. (Linguraru et al., 2012). The liver biopsy process is, however, an invasive process involving chirurgical procedures on patients, and such processes are expensive and time-consuming. A large number of scientific projects have also been carried out to discover and improvise non-invasive methods, along with the assistance of technological advancements.

It should be remembered, however, that the diagnostics of non-invasive approaches have not been recognized internationally yet and the adoption rate differs between countries. Such procedures nevertheless act as pre-screening measures, which allow physicians to reduce the populations of patients that eventually need to be subjected to biopsy. In comparison with traditional biopsies, such diagnostic processes are also quick and, therefore, early detection can lead to early initiation of treatment procedures (Meng et al., 2017). As a result, these medical concerns have gained significant attention in recent years regarding the way to design future liver diagnostics and how to develop personalized treatment plans. The objective is to find a non-invasive alternative for the traditional diagnostic methods that rely primarily upon histopathological analysis and biopsy.

Ultrasonic imaging technologies are one of the most effective forms of evaluation, detection, and treatment of different types of clinical diseases through their major advantages, including real-time functionality, high precision, strong accessibility, and non-invasive assessment, as an outcome of the advancement of ultrasonic imaging technology. Another method is the computerized axial tomography (CT) technology (Himmah, Sigit and Harsono, 2018), which is often used to produce photographs of organs, which cannot be seen by traditional X-ray camera equipment, with an excellent and precise resolution of the resulting image. However, images may contain organs that are not necessary for diagnosis; segmentation is therefore done to solve this problem.

A classification framework is established in this study to identify and distinguish between healthy or normal and unhealthy or abnormal liver. The suggested Computer-aided Diagnostic architecture consists of image pre-processing, segmentation, extraction, and collection of features and classification using different classifiers based on traditional machine learning, deep learning, and transfer learning. The remainder of this paper is structured accordingly. Section II displays the related work on multiple CAD devices used for liver disease identification and diagnosis. The relevant research methodology and approach is presented in Section III. The architecture and design specifications of the various models used are discussed in Section IV. The execution and implementation of the research goals are covered in Section V. Section VI contains information on the results and evaluation of the experiment. The conclusion and future work of the study is ultimately suggested in Section VI.

1.1 Research Question

How effective, accurate and fast is the detection of liver disorders from CT scan images using the deep learning and transfer learning ?

2 Related Work

Identification and detection of liver diseases are expensive, invasive, and time-consuming when done manually through doctors and pathologists. Therefore, much research has been done in his filed to design automated methods of diagnosis that can be reliable, accurate, and cost-effective. A significant number of researches on the detection of different liver diseases have been done using MRI, CT scan, and Ultrasound images. The recent advancements in the area of transfer learning have added to the betterment of such diagnostic systems. Researches that achieved state-of-the-art results with relevant image processing, deep learning, and transfer learning methods are discussed below.

2.1 Image classification using traditional machine learning

In the study (Roy et al., 2019), the image dataset used for the experiment is the collection of CT (Computed Tomography) scan images of lungs. The raw images are pre-processed and the segmentation is carried out in the Region of Interest (ROI). The authors have used the Random Forest method for the classification of distinct features. The objective of this study is to build a model that is capable of classifying the images into two categories namely, healthy and unhealthy. The two traditional machine learning methods used include Random Forest and Support Vector Machine (SVM) and after experimentation, it is found that the SVM classifier performs better, achieving 94.5% efficiency, which is commendable. However, there is a scope of improvement, and improvements can be attempted by selecting a dataset containing more images and using better traditional classification methods or deep learning approaches. In the study (Prajith, Kumar and Kareem, 2016), the image dataset used for the classification problem is the collection of ultrasonic images of the liver. The machine learning methods used for the classification of images into categories include Artificial Neural Networks (ANN), Gaussian Mixture Model (GMM), and Support Vector Machine (SVM). Comparing the three used different models, it is found that the SVM technique outperforms the other two models because it achieves a classification accuracy of 94%, specificity as 95 %, and sensitivity as 93.33 %. These results are appreciable, however, the size of the dataset used is 219 only, which should be further improved in future research and also, better feature extraction techniques must be employed to attain better classification metrics. In the study (Alquran et al., 2017), the objective is to develop a non-invasive image analytics system that can assist in the detection of melanoma skin cancer. The approach used is based on Support Vector Machine (SVM), but a series of steps is applied to the dataset before building the model. The various steps include image pre-processing techniques such as segmentation using thresholding, feature extraction, and feature selection using Gray Level Co-occurrence matrix and Principal component analysis, respectively. After these steps of data pre-processing and augmentation are carried out, data is trained using SVM to obtain a classification accuracy of 92.1%, which is remarkable. In the study (Makaju et al., 2018), the objective is to evaluate the existing techniques and models related to model building and image pre-processing, and finding the one that achieves the best results, and lastly, proposing the model capable of achieving even better results by overcoming the limitations and drawbacks of the existing best model. The image dataset used for the experiment consists of 1018 CT scan images in DICOM format having image size as 512 * 512 pixels. Various pre-processing techniques such as greyscale, median filter, Gaussian filter are used along with segmentation and feature extraction, and 5-fold cross-validation to improve the performance. The proposed model is based on watershed segmentation and SVM classifier, and experimental results show that the classification accuracy of 86.6% is achieved. However, there is a scope of future work by designing the model to classify cancer into different stages to provide more granular information.

2.2 Image pre-processing techniques

In the study (Himmah, Sigit and Harsono, 2018), the objective is to achieve the best possible accuracy using the watershed segmentation algorithm on the abdominal CT scan images to segment the liver from the background. The various processes applied to raw CT scan images include erosion and dilation, then watershed segmentation, then cropping and median filtering. The experimental results show that an average segmentation accuracy of 98.28% is achieved and it can be concluded that the watershed algorithm is suitable for such applications, such as segmenting liver from the abdominal CT scan images. In the study (Midhila, Krishnan and Sudhakar, 2017), the authors present a report on the various techniques carried out in the image processing phase of ultrasound images of the liver. The different techniques discussed include noise removal, segmentation, feature extraction, and classification, and their accuracies are then compared. Moreover, Gray difference weight segmentation is used on different types of liver diseases. Overall, the paper provides a great amount of knowledge for the classification with different types of liver diseases when working with the ultrasound images of the liver. In the study (M. Hassan, Elmogy and Sallam, 2015), The objective is to design a computer-aided diagnostic system for the classification of medical images. The image dataset used consists of 110 liver ultrasound images belonging to three different categories, namely cyst, haemangioma, and hepatocellular carcinoma. Various pre-processing techniques are employed such as noise removal, contrast enhancement, median filter, and segmentation using fuzzy c-means clustering. The model used for the classification multi-support vector machine (multi-SVM) and later, 10-fold cross-validation is employed to improve the system performance. Overall, the system achieves a classification accuracy of 96.5%, which is remarkable. Despite the usage of remarkable techniques, the size of the dataset is not appreciably large and thus, results are not scalable. In the study (Lawankar, Sangewar and Gugulothu, 2016), the objective is to performs the segmentation of the liver from the abdominal CT scan images. In the process, various other steps are also followed, such as noise removal, greyscale, contrast enhancement, and gradient magnitude. Finally, the watershed segmentation algorithm is applied to obtain the region-of-interest (ROI) with an accuracy of 92.1%.

2.3 Image classification using Deep Learning

In the study (Wei et al., 2019), the objective of the experiment is to build a model based on Artificial Neural Network (ANN) in order to find the most crucial predictors involved in non-invasive liver fibrosis reverse. The dataset used for conducting the study involved the relevant laboratory data for 298 patients for a period of almost 2.5 years. Univariate and multivariate analysis of reverse and non-reverse groups is carried out at the baseline and after 1.5 years and the results are compared. Moreover, a logistic regression approach is used. The experimental results show that the ANN model performs better than Logistic regression and the most crucial predictors are found, which might help in improved diagnostic. Overall, the concept of finding the influential predictors is remarkable, however, the dataset size is not satisfactory and results can be made scalable through a larger dataset. In the study (Ker et al., 2019), the objective is to build a model that can classify the CT scan images of the brain into healthy and unhealthy. The approach used for this classification is 3-Dimensional Convolutional Neural Networks. Both 2-class (binary) and 4-class (categorical) classifications are carried out on 399 volumetric CT scan images. The image thresholding technique is used in the image pre-processing step to improve the model accuracy and performance of the overall classification system. The results are compared with and without the image thresholding technique and achieving the highest f1-score of 0.952. In conclusion, the research is remarkable in approach and results and outperforms other works applying 3D-CNN to CT scan and MRI scan images of the brain. In the study (Frid-Adar et al., 2018), the method for creating synthetic medical images is discussed. The approach used is based on Generative Adversarial Network (GAN) and is applied on a dataset of 182 CT scan images of the liver. Firstly, the GAN model is used for data augmentation by producing synthetic images. Secondly, traditional data augmentation techniques are used to increase the size of the dataset. Finally, the model based on CNN is trained and the performance of the model is compared in both the cases, i.e. using classic data augmentation techniques and using synthetic data augmentation based on GAN. The experimental results show that the model trained on images generated synthetically using GAN performs better with 85.7% sensitivity and 92.4% specificity. In the study (Doğantekin, Özyurt, Avcı and Koç, 2019), a novel approach based on a hybrid combination of CNN and the block-based perceptual hash function is proposed. The objectives include minimizing the execution time of the CNN model, reducing the disk space occupied by the liver images, and lastly, establishing the model with classification accuracy better than the conventional models. The image dataset for the study consists of 200 CT scan images equally divided between the two classes, and 5-fold cross-validation is employed as well. The experimental results show that CNN performed better with the Extreme learning machine approach and giving an accuracy of 97.3%, which is remarkable. However, the size of the dataset is relatively small and thus, results are not scalable and may be improved with augmentation techniques in the future. In the study (Ben-Cohen et al., 2018), the objective is to build an automatic system that can detect liver metastases. The methods used include a fully convolutional network and local superpixel sparsity-based classification. The dataset used consists of CT scans from 34 patients comprising of 123 lesions divided into training and testing datasets. Moreover, 3-fold cross-validation is used to improve the experimentation results, and overall, the results achieve a true positive rate (TPR) of 94.6%, which is appreciable. However, it is clearly noticeable that the dataset size is small, and this limitation can be explored in future studies with larger datasets. In the study (Alquran et al., 2017), the objective is to develop a non-invasive image analytics system that can assist in the detection of melanoma skin cancer. The approach used is based on Support Vector Machine (SVM), but a series of steps is applied to the dataset before building the model. The various steps include image pre-processing techniques such as segmentation using thresholding, feature extraction, and feature selection using Gray Level Co-occurrence matrix and Principal component analysis, respectively. After these steps of data pre-processing and augmentation are carried out, data is trained using SVM to obtain a classification accuracy of 92.1%, which is remarkable.

2.4 Image classification using Transfer Learning

In the study (Yu et al., 2018), the objective of the experiment is to perform the classification of images into different categories without the need of manual pre-processing methods such as segmentation and feature extraction. It was made achievable by the use of a deep neural network based on the concept of transfer learning, namely AlexNet-Convolutional Neural Network (CNN). Furthermore, the performance of this model is compared with various models that require manual segmentation and feature extraction, such as artificial neural networks (ANN), support vector machines (SVM), multinomial logistic regression (MLR), and random forests (RF). This study is based on liver fibrosis of rat livers and used about 100 labelled image samples, the pre-processing of images is carried out in MATLAB software and techniques used include greyscale, contrast enhancement and adaptive-thresholding. The results revealed that the accuracies obtained in both approaches, i.e. automated pre-processing and manual pre-processing, is similar. However, it is noticeable that the size of the dataset is not large and the results are not scalable. Nevertheless, this study provides a gateway to a variety of experiments in this field. In the study (Dandan et al., 2019), the objective is to design a classification framework that can classify the ultrasound images into three categories, namely normal, fatty liver and liver fibrosis. The image dataset consists of 2942 ultrasound liver images of the three mentioned categories. The proposed model uses CNN and multi-scale grey-level co-occurrence matrix for extraction of image structure features and extraction of image texture features, respectively. After pre-processing, pre-trained GoogleNet is used to train the model and then, lightGBM classifier is used to classify the image into three categories. Overall, the achieved accuracy is 82.6% which can be further increases to 88% by the use of texture features. In the study (Meng et al., 2017), the objective is to design a novel classification method based on the concept of transfer learning and deep learning. The image dataset used in the research consists of 279 region-of-interest ultrasound images of the liver for three categories, namely, healthy, early-stage fibrosis ad last-stage fibrosis. In case of insufficient samples, deep feature extraction can be achieved through transfer learning models, thus, this research uses VGGNet and later on, the deep features are fed into Fully Connected Network (FCNet) for the classification of images into three mentioned classes. The experimental results show that accurate models can be designed using a combination of deep feature extraction and classification. However, the size of the dataset is not large and thus, can be increased through relevant augmentation techniques. Overall, the combination of models is novel and scalable in future studies. In the study (Reddy, Bharath and Rajalakshmi, 2018), the objective is to design a computer-aided diagnostic system capable of classifying the fatty liver into normal and unhealthy classes. The proposed method is based on a combination of convolutional neural networks and a transfer learning model called VGG-16. The image dataset used for the research consists of 157 ultrasound images of normal and fatty liver disease livers from different age groups and different genders. The experimental results show that the classification accuracy of 90.6% is achieved by the proposed system, which is commendable. However, the dataset size is insufficient, which is a common problem for medical images data collection. Nevertheless, results and approach are appreciable and can be taken forward in future studies with the use of better combinations of models and transfer learning strategies. In the study (Chen et al., 2019), the objective is to solve the problem of spatial information loss faced when traditional convolutional encoder-decoder is used for the segmentation process. The proposed method is NL-Net which consists of three blocks namely, encoder, learning based on non-local spatial information, and decoder. Encoder block uses a pre-trained model called ResNet. The image dataset is from a renowned challenge and comprises of 6407 abdominal CT scan images. The experimental results show that an average dice of 0.972 is achieved, which states that 97.2% segmentation is accurate. The study results are remarkable and future researchers can gain great insights from it. In the study (Balagourouchetty, Pragatheeswaran, Pottakkat and G, 2020), the objective is to design the most accurate computer-aided diagnostic system possible. The approach is based on the combination of transfer learning model GoogleNet and convolutional architecture-based FCNet classifier. Some further modifications are made to enhance the system performance such as replacing ReLu with leaky-ReLu, three additional fully connected layers before classification layer, and feeding the output of each layer as input to next inception layer. The image dataset used consists of CT scan images of liver comprising of six different types of diseases. In conclusion, the proposed method of combining models is novel and results in remarkable classification accuracy. In the study (Dawud, Yurtkan and Oztoprak, 2019), the objective is to design a computer-aided diagnostic system for classification of brain haemorrhage from the CT scan images of the brain. The research uses three different classification approaches and then their performances are compared, including CNN, pre-trained model AlexNet, and a novel combination of AlexNet and SVM called AlexNet-SVM. The images are classified into two categories, containing haemorrhage or not, and various augmentation techniques are used to increase the size of the dataset. All the models are trained using the same dataset of CT scan images and the experimental results show that the AlexNet-Net outperforms the other two approaches by achieving a classification accuracy of 93.48%. This study provides great insights on how slight fine-tuning can lead to better results. In the study (Wang, Dong, Wang and Wang, 2020), the objective is to design a computer-aided diagnostic system for non-invasive detection of lung cancer. The proposed method uses a residual neural network based on medical-to-medical transfer learning. The image dataset consists of CT scan images of lungs and there are two datasets, one is publicly available and another is collected from a hospital and data is confidential. The model is initially trained on a publicly available dataset and then, on the collected dataset after undergoing fine-tuning steps.

Kera’s Image Data Generator is used for the various tasks of loading and augmenting data. Finally, the accuracies of different models trained on the same dataset are compared. The experimental results show that the residual neural network achieved an accuracy of 85.71%, outperforming AlexNet, VGG16 and DenseNet-121.

2.5 Summary

Several problems and issues have been identified from the relevant literature review. The lack of large datasets with appropriate labelling prevents the advanced models from generalizing properly. The time and money needed for the model's preparation and development are immense and this concept can only be evaluated through study and experiments that have access to these resources. There are two methods to use transfer learning in which one involves the introduction of new layers or re-training of the pre-trained model and the other involves using the model without training layers only for the retrieval of the features. This study implements image processing techniques and uses several models such as SVM, a CNN built from scratch, pre-trained DenseNet-169 and Inceptionv4 for classification with parameter tuning. This work will expand the DenseNet-169 paradigm for the transfer-learning used in multiple medical trials but new for liver CT scans. Moreover, the use of Inceptionv4 is entirely novel for liver disease detection.

3 Research Methodology

This research work implements several deep learning models to abdomen CT scan images to classify them into a healthy or normal liver and unhealthy or abnormal liver with the best possible accuracy. The CRISP-DM approach is implemented to ensure that the research systematically achieves its goals. It entails six stages: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment. (Pérez et al., 2015).

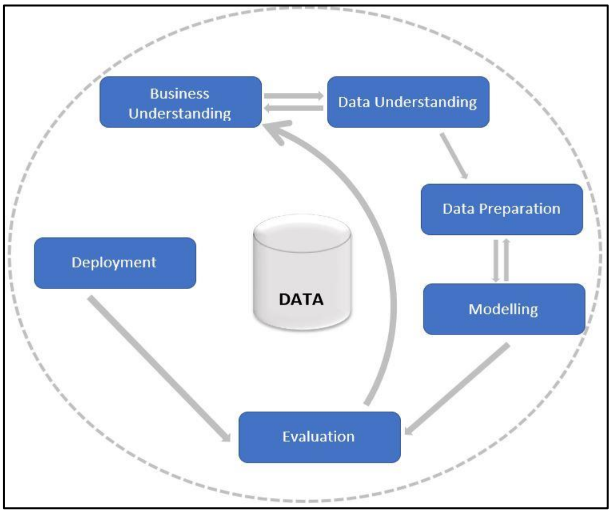

Figure 1 CRISP-DM Life Cycle

3.1 Business Understanding

Ultrasound, CT scans and MRI images are popular techniques for the investigation and treatment of diseases of the liver, but biopsy produces results with incomparable accuracy.

Biopsies are nevertheless invasive, costly, time-consuming and require specialists. Moreover, it is understandable that the evaluation of biopsy images requires various resources from pathologists and is heavily dependent on the medical professionals. This leads to the requirement of developing non-invasive techniques for the diagnosis where the images labelled by pathologists and doctors are used as the base and are pre-processed to gain the maximum amount of details possible. This minimizes the involvement of professionals, is less time-consuming and less prone to human errors, and can be utilized as a pre-screening test. This research utilizes an SVM classifier, deep learning-based CNN model and transfer learning-based models to reduce the time required for evaluating CT scan liver images. The transfer learning models DenseNet-169 and Inception-v4 are used to achieve high accuracy with an insufficient number of samples.

3.2 Data Understanding

This is difficult and expensive to procure data sets containing actual patients' diagnostic image samples rather than digitally generated samples. This work gathers 200 pictures of two separate classes, i.e. healthy or normal liver and unhealthy or abnormal liver, using an image source. This image source is MedPix[1], a free open-access archive of digital photographs that medical schools, medical practitioners and academics can use. This provides a wide variety of Liver CT scans to conduct this research. Originally developed by the Departments of the Radiology and Biomedical Informatics of Uniformed Services University, Bethesda, Maryland, USA, the National Library of Medicine is a free electronic reference image archive. Table 1 explains the image distribution into two classes. All the images are three-dimensional of various sizes and are processed in the following steps. The data preparation steps are addressed in the following section, considering all the information relating to image data.

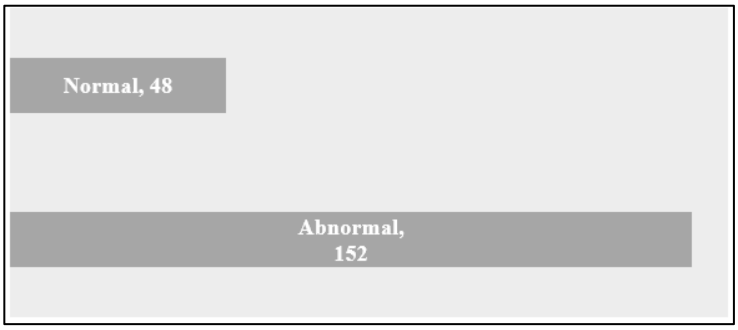

Figure 2 Data Class Distribution

3.3 Data Preparation

The dataset consists of 200 CT scan images belonging to two classes. Some of the downloaded images were in different formats; therefore, all the images types were transformed to ".jpg" to maintain consistency and uniformity of data. The approach followed in this research splits the data into training and testing datasets such 80% of the total images are used for training and the remaining 20% for testing the model. Furthermore, the training dataset is divided into two parts such that validation dataset has 20% images and remaining images forms the final training dataset. Images are pre-processed to capture full information and functionality before feeding

[1] The URL is https://medpix.nlm.nih.gov/.

them to CNN and DenseNets-169 models. Not just that, but images are then augmented through different techniques to up-sample the size of the image dataset. Runtime augmentation further increases the number of image samples using various parameters such as rotation, zoom, scaling, batch size and size of the image. These procedures are covered in depth below.

3.3.1 Image Pre-processing

The image repository has been generated to represent each directory as a label/class with the images in < image id > syntax. The first task was then to group all images into healthy or normal and unhealthy or abnormal categories. The root directory was then split into separate train and test directories with a stratified ratio of 80/20. The train folder was then assigned with 160 images and 40 images were transferred to the test folder. The data are moved to the cloud instance (Google Drive) only once the root directory is pre-processed, cleaned and split, to order to reduce the cost of data transmission to the cloud. All images from the root directory to the child directories have been loaded into the models using ImageDataGenerator class of the Keras library. As defined in the following section, it also transforms and increases the number of images. The ImageDataGenerator class parameter is set as binary as two classes classify the results. Histogram Image equalization, based on average and variance is executed to improve the contrast and brightness of the image without influencing the specifics of image data. Besides, the images are segmented using watershed transform to render liver positions that can discern artefacts by context and separate an image into liver parenchyma and non-parenchyma areas of the liver. It is an efficient region-based grayscale image segmentation algorithm when the two regions-of-interest converge on the edges. The image is viewed as a topographic map, with a height reflecting the size of each pixel.

3.3.2 Image Transformation and Augmentation

Diverse data transformations are used to ensure that the model does not overfit the training data and to assure that the test set has reasonable validation accuracy. The subsequent methods were used to subdue certain complications after diligent analysis and methods such as image augmentation, batch normalization, and dropout layers were utilised. Various techniques implemented are as follows:

- Rescale/Resize: Scaling is the image re-dimension. The image size may be manually defined or scaled by a factor.
- Zoom: It enlarges the image spontaneously, making feature information clearer.
- Width Shift: It shifts the image on the horizontal axis.
- Height Shift: It shifts the image on the vertical axis.
- Rotation: Image is rotated according to defined angle degree.
- Flip: It flips the image horizontally and vertically as specified.
- Fill mode: After transformation, it completes the newly formed pixels in the image.
- Crop: The image composition is modified by eliminating the undesirable regions in the image.
- Gaussian Blur: It is a technique to blur an image with a Gaussian function to lower the background noise and to minimize detail.
- Greyscale: It transforms RGB-resolution colour images into grey images where each pixel either reflects light or shadow in terms of intensity.
- Averaging Blur: This is a method of smoothing which lowers the difference in intensities seen between neighbouring pixels.
- Median Blur: It is a non-linear technique used to remove noise from an image.

- Bilateral Blur: It is a non-linear filter for images, which protects the edges and reduces noise simultaneously.
- Erosion: It is useful for removing small white noises by eroding away the boundaries of the foreground object.
- Dilation: It is just opposite of erosion and it increases the white region in the image or size of the foreground object. It follows erosion to increase objects size after removing noise.
- Morphological Gradient: It is a technique based on the difference between dilation and erosion of an image and results in an outline of the object.
- Sharpen: It is a technique for increasing the sharpness of an image and makes the edges clearer.
- Adaptive Gaussian Noise: It adds a statistical noise having a probability density function equal to normal distribution.
- Contrast: This method clarifies the image's characteristics by maximizing the use of colours, and adjusts the spectrum in an image to improve contrast.
- Affine Transformation: This technique transforms the image keeping the parallel lines in the image structure intact.
- Histogram Equalisation: It is an image processing technique used to improve contrast in images.
- Watershed transform: It segments the regions that touch each other.

3.4 Modelling

This section explains the architecture used for the SVM, CNN, DenseNet-169 and Inception v4 models.

3.4.1 Traditional Machine Learning using Support Vector Machine (SVM)

The SVM is a supervised classification and regression algorithm that uses algorithms and SVM kernels to analyse the data. Supporting vector classification (SVCs) is also an algorithm which seeks an optimum surface separation. When complete separation of the two classes is not possible, SVM kernel methods are used. The polynomial, quadratic and radial basis function are types of kernels of SVM (Malek et al., 2019). Any data object in the SVM algorithm is drawn as a point in n-dimensional space, where n is the number of features, so each feature's value is the value of a given co-ordinate. Classification is then carried out by determining the hyperplane that optimally separates the two groups. Model efficiency can be improved by the fine-tuning of parameters like C, Gamma and Kernel.

3.4.2 Deep Learning using Convolutional Neural Network built from scratch

CNN consists of a series of neural networks for the extraction of features and concentrates all features from previous layers on the various levels of the system. The architecture of CNN consists primarily of five levels (Valueva et al., 2020). The convolution layer is used to obtain spatial knowledge from the given input. The filter measures the weight and image area. The batch normalisation layer checks the output from the previous layer and sends it to the next level of activation. Because normalization is applied during the gradient process, weights in the network are balanced. The Rectified Linear Unit (ReLU) is used to trigger the CNN. As CNN deepens, the complexity of the features is increasing. MaxPooling layer is being used to minimize the size of the map of the function. For category count calculations each fully connected layer neurone binds to the previous layer. The output layer is comprised of a classifying function softmax or sigmoid. The architecture comprised of 3 convolution layers with a filter size 3x3, 3 max pooling layers of 2x2, 1 flat layer which converted the data into one-dimensional array, one dense layer with 512 neurons and L2 regularizer, and lastly, dense layer for classification with 2 neurons and sigmoid function.

3.4.3 Transfer Learning using pre-trained Inception-v4 model

Inception v4, which has been developed by Google and is an advanced version of Google Net, was another model for transfer learning used in this study. Due to its compact architecture and more inception blocks, it is better able to achieve results than Inception V3 and ResNet versions. In comparison, there are important effects of merging inception networks with residual networks. It provides support for batch normalization, memory optimisation and backpropagation (Emara, Afify, Ismail and Hassanien, 2019).

3.4.4 Transfer Learning using pre-trained DenseNet-169 model

Transfer learning is the concept of knowledge transfer wherein the knowledge gained through training on one type of data is used on another type of data. Mostly, such an approach is employed the size of the original dataset is small. The transfer learning approach used in this research study is DenseNet-169. DenseNet model is a logical extension of ResNet models and proves to be more efficient. The model uses 169 layers and pre-trained weights from Imagenet. The rationale behind using transfer learning is that such models perform appreciably on images with a less computational cost even when parameters are less (Huang, Liu, Van Der Maaten and Weinberger, 2017).

4 Design Specification

The design and architecture of each model are discussed in this portion of the paper. Deep learning and transfer models are made up of different layer types that allow the model to train and retrieve features. These are the various layers used in the model structure.

- Convolution layer: It is the layer with several filters to extract features depending on the kernel size. The number of filters is the same as the rate of growth for extraction functions in this layer (Khened, Kollerathu and Krishnamurthi, 2019).
- Batch Normalisation: This layer is used to standardize the performance of one convolution to another convolution, and thereby, reduces the overfitting problem.
- Max Pooling: Max pooling layer is used by taking the highest value from the kernel size of a certain area to reduce the dimensionality of the feature map.
- GlobalAveragePooling2D: The average value of the feature maps reduces the dimension to one.
- Dense (Fully Connected): These are the layers normally positioned before the classification layer and incorporates information derived from the feature maps of previous layers.
- Dropout layer: This layer decreases overfitting when the percentage of features in the model is lowered. 0.5 Dropout is defined as a parameter.
- Softmax layer: This is the last dense layer with activation function as softmax and it generates the final classification label map, generally used for multi-class classification.
- Sigmoid layer: This is the last dense layer with activation function as sigmoid and it generates the final classification label map, generally used for binary classification.

- Loss Function: The cross-entropy loss function is employed to resolve the disparity in class between background and region-of-interest.
- Weight: In the transfer learning, instead of using 'imagenet' pre-trained weights, different weights are used for training DenseNet-169 without including the top layer.
- Optimizer: Adam optimizer with a learning rate of 0.0001 or 1e-4 is employed.

5 Implementation

This section provides information on how the different models are implemented to classify the liver CT scan images into normal or healthy and abnormal or unhealthy categories.

5.1 Environment Setup

CNN and transfer learning take longer time and more memory to process the images during model preparation. Therefore, the experiment is carried out on the Google Colaboratory with 15 GB drive storage, 12.72 GB RAM, 107.77 GB disk space, and run-time on-demand GPU support. Google Colaboratory notebooks are just like Jupyter notebooks but hosted on the cloud rather than the local machine. These notebooks use python version 3.6.9, and libraries such as OpenCV, Keras and TensorFlow are employed for image pre-processing and execution of CNN, DenseNet169 and Inception-v4 models. The time a model takes to learn depends on the number of layers involved, and thus, runtime GPU support is beneficial as it further expedites the execution time of these models. All data is uploaded on Google Drive and accessed in notebooks by mounting Google Drive using python library available for this purpose. Image pre-processing, transformations and, up-sampling are done using cv2 (OpenCV) and Keras libraries in python.

5.2 Data Handling

All CT scan images are read and loaded into the directories from Google Drive. Different directories were made according to train, test and validation data belonging to normal and abnormal classes. The original 200 images were used initially to train the model, but the results were inadequate. Therefore, to overcome this issue, images were augmented using OpenCV cv2 library and all the images were stored in Google Drive. Furthermore, images were split into three folders namely, Training, Validation and Testing, containing images of both the categories. ImagedataGenerator class from the Keras library is further used for up-sampling the datasets during run-time. Various parameters are used and modified to improve the overall model performance. ImageDataGenerator class provides a function called flow_from_directory that allows loading data into train, valid and test data generators with appropriate parameters.

5.3 Architecture

All the libraries and packages required to implement TensorFlow using Keras were imported into the Colab notebooks and uninstalled packages were installed using pip command. After setting up the environment with all needed libraries and the mounted drive containing pre-processing data, it was time to build the architecture of the different models. The 10-fold cross-validation methodology is used in models to guarantee that the model does not overfit and the variance tends to be minimal.

5.3.1 Support Vector Machine (SVM)

The initial model is the traditional machine learning model, Support Vector Machine (SVM) because SVM has been successful in image classification tasks as seen in Section 2. The SVM model is implemented with GridSearchCV hyper-parameter tuning technique to optimise the model performance.

Table 1 SVM Parameters

Parameter	Value
Cost (C)	1, 10, 100, 1000
Gamma	0.001, 0.0001
Kernel	Linear, Radial basis function (rbf)

5.3.2 Convolution Neural Network

The next model is the Convolution Neural Network based on deep learning. The architecture is built using 3 convolution layers of filter size of 16, 32 and 64 with kernel size (3,3). The first convolution layer requires the input shape, and the activation function used in every convolution layer is Rectified Linear Unit (ReLu). After each convolution layer, there is a Max Pooling layer with pool size (2,2) to down-sample the input representation by taking maximum value defined by pool size. Next layer is the Flatten layer that converts or flattens the input to the one-dimensional space without affecting the batch size. Then, a dense layer with 512 neuron units, activation function ReLu and L2 regularizer, is added to the sequential model. Finally, the classification layer is built using Dense layer with 2 units Sigmoid activation function is added to perform the binary classification. The model is compiled with Adam optimiser having a learning rate of 0.0001 and loss function 'binary_crossentropy'. Moreover, 'Early Stopping' and 'Reduce LR on Plateau' classes are used to monitor and regulate model performance.

Table 2 CNN Model Architecture

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 297, 297, 16)	448
max_pooling2d (MaxPooling2D)	(None, 148, 148, 16)	0
conv2d_1 (Conv2D)	(None, 146, 146, 32)	4640
max_pooling2d_1 (MaxPooling2	(None, 73, 73, 32)	0
conv2d_2 (Conv2D)	(None, 71, 71, 64)	18496
max_pooling2d_2 (MaxPooling2	(None, 35, 35, 64)	0
flatten (Flatten)	(None, 78400)	0
dense (Dense)	(None, 512)	40141312
dense_1 (Dense)	(None, 2)	1026
Total params: 40,165,922		
Trainable params: 40,165,922		
Non-trainable params: 0		

Table 3 CNN Model Parameters

Parameter	Value
Batch Size	16, 32
Epochs	10
Optimiser	Adam
Learning Rate	0.0001
Early Stopping Monitor	Validation Accuracy
Early Stopping Patience	2
Reduce LR on Plateau Monitor	Validation Loss

5.3.3 Inception-v4

The Inception-v4 model is an extension model of GoogleNet and Inception series with advanced inception and reduction blocks. This study uses Inception-v4 with 'inception-v4_weights_tf_dim_ordering_tf_kernels_notop.h5' pre-trained weights. The top classification layer is not included and fully connected architecture is built instead of the top layer. A Global Average Pooling layer is added, followed by Batch Normalisation layer. Then, Dense layer with 1024 units and 2048 units and ReLu activations are added to the model with Dropout layers with a dropout rate of 0.5 between them. Finally, a Dense layer with 2 neurons and softmax activation function to perform the binary classification. The model is then compiled with Adam optimiser having learning rate 0.0001 and loss function ‘binary_crossentropy’. Furthermore, ‘Early Stopping’ and ‘Reduce LR on Plateau’ classes are used to monitor and regulate model performance.

Table 4 Inception-v4 Model Parameters

Parameter	Value
Weights	inception-v4_weights_tf_dim_ordering_tf_kernels_notop.h5
Batch Size	16, 32
Epochs	10
Optimiser	Adam
Learning Rate	0.0001
Early Stopping Monitor	Validation Accuracy
Early Stopping Patience	2
Reduce LR on Plateau Monitor	Validation Loss

5.3.4 DenseNet-169

The DenseNet-169 model is a DenseNet model with 169 layers. This study used it with 'DenseNet-BC-169-32-no-top' pre-trained weights. The top classification layer is not included and fully connected architecture is built instead of the top layer. A Global Average Pooling layer is added after DenseNet, followed by Dropout layer with a dropout rate of 0.5, and finally, a Dense layer with 2 neurons and sigmoid activation function to perform the binary classification. The model is then compiled with Adam optimiser having learning rate 0.000005 and loss function ‘binary_crossentropy’. Furthermore, ‘Early Stopping’ and ‘Reduce LR on Plateau’ classes are used to monitor and regulate model performance.

Table 5 DenseNet-169 Model Parameters

Parameter	Value
Weights	DenseNet-BC-169-32-no-top.h5
Batch Size	16, 32
Epochs	10
Optimiser	Adam
Learning Rate	0.000005
Early Stopping Monitor	Validation Accuracy
Early Stopping Patience	2
Reduce LR on Plateau Monitor	Validation Loss

Table 6 DenseNet-169 Model Architecture

```
Layer (type)                 Output Shape              Param #
=================================================================
densenet169 (Functional)     (None, 9, 9, 1664)        12642880
_________________________________________________________________
global_average_pooling2d (Gl (None, 1664)              0
_________________________________________________________________
dropout (Dropout)            (None, 1664)              0
_________________________________________________________________
dense (Dense)                (None, 2)                 3330
=================================================================
Total params: 12,646,210
Trainable params: 12,487,810
Non-trainable params: 158,400
```

6 Evaluation

This segment of the report examines the models and all the parameters fine-tuned to achieve the optimal results. Four models are utilised for this research on liver CT scan images. Models such as DenseNet-169 and Inception-v4 are applied to make the research novel and innovative. Initially, the training and validation accuracies and losses calculated for each epoch, are used to evaluate the model performance. This can be analysed through plots which reveals how the accuracy and loss of training and validation datasets vary with each epoch. Moreover, the test accuracy is calculated through predictions because test data is completely new for the model. Additionally, confusion metrics and classification reports are generated for each model to gain acumens such as precision, recall, f1-score, true positives, true negatives, false positives and false negatives. Each model is evaluated in a separate experiment subsequently.

6.1 Experiment 1: Support Vector Machine

The first model used in this work was the Support Vector Machine (SVM). SVM is preferred because image classification is a challenging task, and a literature review explicates that SVM succeeds. In this algorithm, the data points are categorized into classes by locating a hyperplane in the N-dimensional space, which acutely classifies the data points. This model is applied with GridSearchCV technique of hyper-parameter tuning to obtain the optimal model performance. The model performance is evaluated through the classification report that can be generated using metrics library from sklearn package. It can be observed that the model achieves an accuracy of 99%.

```
Classification report for -
GridSearchCV(cv=None, error_score=nan,
             estimator=SVC(C=1.0, break_ties=False, cache_size=200,
                           class_weight=None, coef0=0.0,
                           decision_function_shape='ovr', degree=3,
                           gamma='scale', kernel='rbf', max_iter=-1,
                           probability=False, random_state=None, shrinking=True,
                           tol=0.001, verbose=False),
             iid='deprecated', n_jobs=None,
             param_grid=[{'C': [1, 10, 100, 1000], 'kernel': ['linear']},
                         {'C': [1, 10, 100, 1000], 'gamma': [0.001, 0.0001],
                          'kernel': ['rbf']}],
             pre_dispatch='2*n_jobs', refit=True, return_train_score=False,
             scoring=None, verbose=0):
              precision    recall  f1-score   support

           0       0.99      1.00      1.00       449
           1       1.00      0.96      0.98       114

    accuracy                           0.99       563
   macro avg       1.00      0.98      0.99       563
weighted avg       0.99      0.99      0.99       563
```

Figure 3 SVM Classification Report

6.2 Experiment 2: Convolution Neural Network built from scratch

The following model was constructed from scratch with 3 convolution 2D layers of 16, 32 and 64 filter sizes. The model was executed for 10 epochs and resulted in training accuracy of 98.8% and validation accuracy of 97.95%. Besides, the model variance is tested with 10-fold cross-validation. The accuracy and loss plots for the model are plotted to see the variations for training and validation datasets. Additionally, the model is evaluated using ‘evaluare_generator’ and ‘predict_generator’ functions to check the model performance on unseen test data. Classification report and confusion metrics are generated as well. The model took 1459 seconds to train on the training dataset.

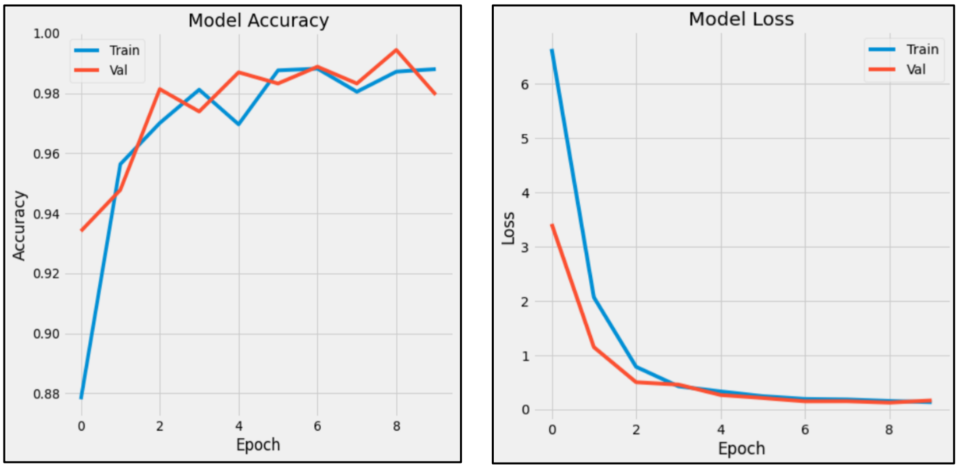

Figure 4 CNN Model Accuracy and Loss

6.3 Experiment 3: Transfer Learning using Inception-v4

This model was employed to implement transfer learning approach. The model build using the pre-trained weights and without including the top layer. The model was executed for 10 epochs but, it stopped after 3 epochs due to early stopping as the validation accuracy was not increasing anymore. It resulted in testing accuracy of 99.83% and validation accuracy of 81.60%. However, the validation loss increased in the following epochs. Besides, the model variance is tested with 10-fold cross-validation. The accuracy and loss plots for the model are plotted to see the variations for training and validation datasets. Additionally, the model is evaluated using ‘evaluare_generator’ and ‘predict_generator’ functions to check the model performance on unseen test data. Classification report and confusion metrics are generated as well. The model took 6655 seconds to train on the training dataset.

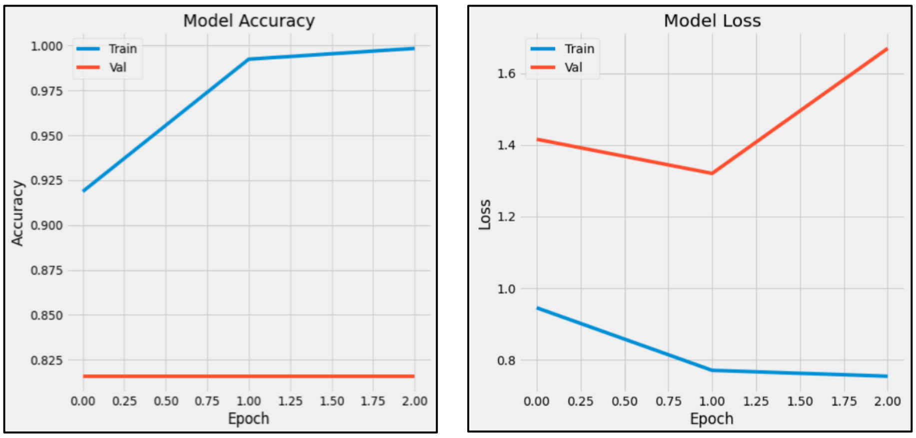

Figure 5 Inception-v4 Model Accuracy and Loss

6.4 Experiment 4: Transfer Learning using DenseNet-169

The last transfer learning model was DenseNet-169 that has 169 layers in its architecture. The model is built using the pre-trained weights and excluding the top layer. The model was executed for 10 epochs but, it stopped after 8 epochs due to early stopping as the validation accuracy was not increasing anymore. It resulted in testing accuracy of 99.66% and validation accuracy of 98.96%. Moroever, the validation loss continuously decreased in every following epoch. Besides, the model variance is tested with 10-fold cross-validation. The accuracy and loss plots for the model are plotted to see the variations for training and validation datasets. Additionally, the model is evaluated using ‘evaluare_generator’ and ‘predict_generator’ functions to check the model performance on unseen test data. Classification report and confusion metrics are generated as well. The model took 13906 seconds to train on the training dataset.

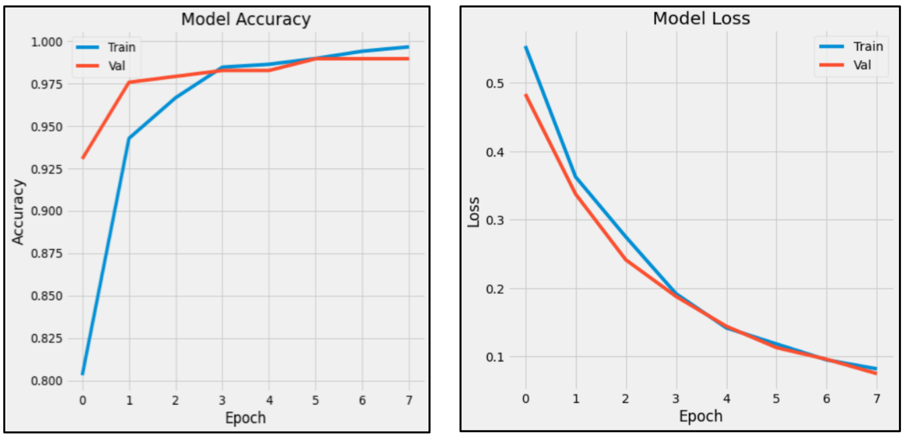

Figure 6 DenseNet-169 Model Accuracy and Loss

6.5 Experiment 4: Computation time of models

After completing the execution of different experiments, an analysis of training times required by the different models was done. The computation time is the execution time of a function or a program and is measured in seconds. When the execution time involved is exceptionally large, then GPU support is essential as it significantly reduces the computational time. A comparison of computational time by different models is carried out and it can be observed that the CNN model proves to be the fastest among all the others.

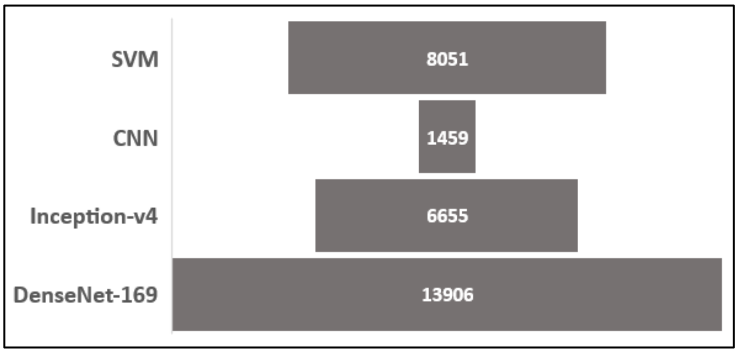

Figure 7 Computational Time Comparison

6.6 Discussion

The key aim of the research was to build an accurate, precise and faster model using deep learning and pre-trained models based on transfer learning. The process of image pre-processing and augmentation plays a pivotal role in the performance of all the models employed. Moreover, 10-fold cross-validation to test the variance in results is equally essential to improve the robustness of the models. The experiments start with the SVM model which attains an accuracy of 99% with recall and f1-score as 0.96 and 0.98, respectively. However, the execution time of SVM is quite high, but it is understandable as it is a traditional machine learning algorithm. The next experiment uses CNN model which is built from scratch and hyper-parameter tuning is implemented to increase the model performance. As a result, an accuracy of 98.80% is achieved on the training dataset and an accuracy of 72% is achieved on testing or unseen dataset with 0.82 recall and 0.83 f1-score in much lesser execution time. Thereafter, first transfer learning model, Inception-v4 was applied and a validation accuracy of 81.60% was achieved. The validation loss starts increasing after some epochs, therefore early stopping cancels the model execution. The last transfer learning model used is DenseNet-169 which is trained in 10 epochs and testing accuracy of 99.66% and validation accuracy of 98.96% is achieved. Moreover, both loss and valid loss continuously decreases with each epoch execution. However, it is worth noticing that DenseNet-169 achieves these results with the highest computation time. As compared to the state-of-art accuracy of 97.3% reported in (Doğantekin, Özyurt, Avcı and Koç, 2019), both CNN and DenseNet-169 models achieved higher accuracy. The reason behind this may be that the state-of-art does not use any image augmentation technique, while this research augments the image data. Overall, CNN and DenseNet-169 models provide the optimal results and are very close, but CNN achieves these results with much lesser execution time.

SVM

	precision	recall	f1-score	support
0	0.99	1.00	1.00	449
1	1.00	0.96	0.98	114
accuracy			0.99	563

CNN

	precision	recall	f1-score	support
Abnormal	0.84	0.82	0.83	246
Normal	0.25	0.28	0.26	54
accuracy			0.72	300

Confusion Matrix

```
[[201  45]
 [ 39  15]]
```

Figure 8 SVM and CNN Metrics Comparison

7 Conclusion and Future Work

This research started with the aim to develop a computer-aided diagnosis system for the identification and diagnosis of liver diseases. The process of liver biopsy is invasive and expensive; therefore, such a computer-aided diagnostic system can be used to overcome these issues. This is implemented through different types of models based on traditional machine learning, deep learning and transfer learning. The models and their architectures are selected on the basis of knowledge gained through rigorous literature review about their performance and computational cost. First of all, the collected data was explored to analyse the various issues such as class imbalance and missing labels. After ensuring that data was correct, a series of image pre-processing and augmentation steps were followed to transform the images and increase the size of the dataset. Finally, the data was split into training, validation and testing datasets using a decided ratio. The four models were built accordingly, such as SVM from python library, CNN built from scratch, Inception-v4 and DenseNet-169 using pre-trained weights. The models were then trained on the overall dataset of 1876 abdomen CT scan images. The CNN model and DenseNet-169 model achieve accuracy of 98.80% and 99.66%, respectively when model execution runs for 10 epochs. While the models accomplished the research goals, it is not without drawbacks. The testing of the model was conducted on the unseen dataset, but it belonged to the same original dataset which was split into training, validation and testing datasets. The technique of 10-fold cross-validation was applied to control the variance, nevertheless, it does not guarantee that model is robust. Therefore, there are several experiments that can be implemented in the future. Firstly, General Adversarial Networks (GAN) can be implemented in future studies to create artificial images and overcome data limitations. Secondly, extensive hyper-parameter tuning can be implemented to further improve model performance. Thirdly, different image enhancement techniques can be employed to further improve the quality of images and improve the count of True Positives (TP) and True Negatives (TN), and reduce the count of False Positives (FP) and False Negatives (FN), because FP and FN are intolerable in the medical field. Fourthly, due to the lack of properly labelled CT scan images data, multi-class classification was not implemented. Thus, in future, hospitals and pathology laboratories can be contacted to collect the multi-class dataset with appropriate and ethical data procedures. Lastly, the computer-diagnosis framework can be deployed on a desktop-based, mobile-based or web-based application so that the task of diagnosis can be seamless and distributed.

8 Acknowledgement

I would like to express my sincere gratitude to my supervisor Mr Hicham Rifai for his valuable guidance and kind supervision throughout the journey. He guided me in every phase of the research and encouraged me to the highest peak. I would also like to thank my parents and friends for their endless assistance.

References

Alquran, H., Abu Qasmieh, I., Mohammad Alqudah, A., Alhammouri, S., Alawneh, E., Abughazaleh, A. and Hasayen, F., 2017. The melanoma skin cancer detection and classification using support vector machine. In: *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. IEEE.

Balagourouchetty, L., Pragatheeswaran, J., Pottakkat, B. and G, R., 2020. GoogLeNet-Based Ensemble FCNet Classifier for Focal Liver Lesion Diagnosis. *IEEE Journal of Biomedical and Health Informatics*, 24(6), pp.1686-1694.

Ben-Cohen, A., Klang, E., Kerpel, A., Konen, E., Amitai, M. and Greenspan, H., 2018. Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations. *Neurocomputing*, 275, pp.1585-1594.

Chen, L., Song, H., Li, Q., Cui, Y., Yang, J. and Hu, X., 2019. Liver Segmentation in CT Images Using a Non-Local Fully Convolutional Neural Network. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE.

Dandan, L., Huanhuan, M., Xiang, L., Yu, J., Jing, J. and Yi, S., 2019. Classification of diffuse liver diseases based on ultrasound images with multimodal features. In: *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE.

Dawud, A., Yurtkan, K. and Oztoprak, H., 2019. Application of Deep Learning in Neuroradiology: Brain Haemorrhage Classification Using Transfer Learning. *Computational Intelligence and Neuroscience*, 2019, pp.1-12.

Doğantekin, A., Özyurt, F., Avcı, E. and Koç, M., 2019. A novel approach for liver image classification: PH-C-ELM. *Measurement*, 137, pp.332-338.

Emara, T., Afify, H., Ismail, F. and Hassanien, A., 2019. A Modified Inception-v4 for Imbalanced Skin Cancer Classification Dataset. In: *2019 14th International Conference on Computer Engineering and Systems (ICCES)*. IEEE.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H., 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321, pp.321-331.

Himmah, F., Sigit, R. and Harsono, T., 2018. Segmentation of Liver using Abdominal CT Scan to Detection Liver Desease Area. In: *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*. IEEE.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K., 2017. Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,.

Ker, J., Singh, S., Bai, Y., Rao, J., Lim, T. and Wang, L., 2019. Image Thresholding Improves 3-Dimensional Convolutional Neural Network Diagnosis of Different Acute Brain Hemorrhages on Computed Tomography Scans. *Sensors*, 19(9), p.2167.

Khened, M., Kollerathu, V. and Krishnamurthi, G., 2019. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical Image Analysis*, 51, pp.21-45.

Lawankar, M., Sangewar, S. and Gugulothu, S., 2016. Segmentation of liver using marker watershed transform algorithm for CT scan images. In: *2016 International Conference on Communication and Signal Processing (ICCSP)*. IEEE.

Linguraru, M., Richbourg, W., Liu, J., Watt, J., Pamulapati, V., Wang, S. and Summers, R., 2012. Tumor Burden Analysis on Computed Tomography by Automated Liver and Tumor Segmentation. *IEEE Transactions on Medical Imaging*, 31(10), pp.1965-1976.

M. Hassan, T., Elmogy, M. and Sallam, E., 2015. A classification framework for diagnosis of focal liver diseases. In: *2015 Tenth International Conference on Computer Engineering & Systems (ICCES)*. IEEE.

Makaju, S., Prasad, P., Alsadoon, A., Singh, A. and Elchouemi, A., 2018. Lung Cancer Detection using CT Scan Images. *Procedia Computer Science*, 125, pp.107-114.

Malek, S., Hui, C., Aziida, N., Cheen, S., Toh, S. and Milow, P., 2019. Ecosystem Monitoring Through Predictive Modeling. *Encyclopedia of Bioinformatics and Computational Biology*, pp.1-8.

Meng, D., Zhang, L., Cao, G., Cao, W., Zhang, G. and Hu, B., 2017. Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. *IEEE Access*, pp.1-1.

Midhila, M., Krishnan, K. and Sudhakar, R., 2017. A study of the phases of classification of liver diseases from ultrasound images and gray level difference weights based segmentation. In: *2017 International Conference on Communication and Signal Processing (ICCSP)*. IEEE.

Pérez, J., Iturbide, E., Olivares, V., Hidalgo, M., Martínez, A. and Almanza, N., 2015. A Data Preparation Methodology in Data Mining Applied to Mortality Population Databases. *Journal of Medical Systems*, 39(11).

Prajith, C., Kumar, A. and Kareem, H., 2016. Supervised classification and prediction of fibrosis seriousness using ultrasonic images. In: *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. IEEE.

Reddy, D., Bharath, R. and Rajalakshmi, P., 2018. A Novel Computer-Aided Diagnosis Framework Using Deep Learning for Classification of Fatty Liver Disease in Ultrasound Imaging. In: *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE.

Roy, K., Chaudhury, S., Burman, M., Ganguly, A., Dutta, C., Banik, S. and Banik, R., 2019. A Comparative study of Lung Cancer detection using supervised neural network. *2019 International Conference on Opto-Electronics and Applied Optics (Optronix)*,.

Valueva, M., Nagornov, N., Lyakhov, P., Valuev, G. and Chervyakov, N., 2020. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathematics and Computers in Simulation*, 177, pp.232-243.

Wang, S., Dong, L., Wang, X. and Wang, X., 2020. Classification of pathological types of lung cancer from CT images by deep residual neural networks with transfer learning strategy. *Open Medicine*, 15(1), pp.190-197.

Wei, W., Wu, X., Zhou, J., Sun, Y., Kong, Y. and Yang, X., 2019. Noninvasive Evaluation of Liver Fibrosis Reverse Using Artificial Neural Network Model for Chronic Hepatitis B Patients. *Computational and Mathematical Methods in Medicine*, 2019, pp.1-8.

Yu, Y., Wang, J., Ng, C., Ma, Y., Mo, S., Fong, E., Xing, J., Song, Z., Xie, Y., Si, K., Wee, A., Welsch, R., So, P. and Yu, H., 2018. Deep learning enables automated scoring of liver fibrosis stages. *Scientific Reports*, 8(1).