Comparative Analysis of 2D and 3D Convolutional Neural Networks for Medical Ultrasound Image Classification

Angelin Beulah S and Sivagami M*

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India
Email: angelinbeulah.s@vit.ac.in (A.B.S); msivagami@vit.ac.in (S.M)
*Corresponding author

***Abstract*—Image processing in the field of modern health care has obtained a significant impact among the medical practitioners in aiding early diagnosis of diseases, treatment planning and disease monitoring. Neural Networks has been showing impressive success in many image analysis tasks. Convolutional Neural Networks (CNNs) framework has been exploited abundantly in medical image processing, in which the choice between two-Dimensional (2D) and three-Dimensional (3D) CNN architectures remains an open question in medical imaging applications. This research has investigated the difference in performance analysis between two-dimensional and three-dimensional CNNs for medical image classification tasks, focusing on accuracy, model complexity, and computational efficiency. Three-dimensional volumetric data of medical ultrasound images had been obtained, pre-processed using appropriate techniques and was given as input to CNN's. Results indicated that 3D CNNs outperformed 2D CNNs in terms of classification accuracy, especially when the image data contained rich 3D spatial information. The 3D CNNs were able to capture spatial relationships and patterns across multiple slices, providing superior feature representations for medical image analysis. The architecture of 3D CNN was further enhanced by adding convolution layers so that it can read the patterns and voxel relationship with better accuracy than the traditional methods.**

***Keywords*—2D-convoultional neural network, 3D-convoultional neural network, 2D-image representation, 3D-image representation, medical imaging, ultrasound image processing**

I. Introduction

The Healthcare industry is the most significant industry in our country that offers care to millions of citizens and indirectly contributes to well-being of the society. Information technology is revolutionizing the healthcare industry by lending a helping hand and artificial intelligence has become tremendous transformational force in a variety of ways that are beneficial to the health of the society and has changed the healthcare industry in recent years. The volume of data in the medical field has been tremendously increasing at a swift rate, which is where the artificial intelligence techniques are employed to handle the massive data, analyse and to perform automation on the diagnosis processes. The clinical decision-making methods of Artificial Intelligence (AI) are more advantageous over the traditional methods for all types of diseases like chronic diseases cancer radiology to analysing and managing the risk in health care. Learning algorithms have dominated the image processing domain with their accurate prediction as they learn from the hidden patterns of the training data of the patients and diagnose the future outcomes [1]. The Fig. 1 shows the relationship between Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) algorithms where AI is the umbrella of science that acts as a term for algorithms that mimic the human abilities. Machine Learning (ML) is the subset of Artificial Intelligence which handles statistical techniques giving the ability to learn from data and Deep learning algorithms are a subset of ML which are state-of--art performance in many tasks with enough data and computational power.

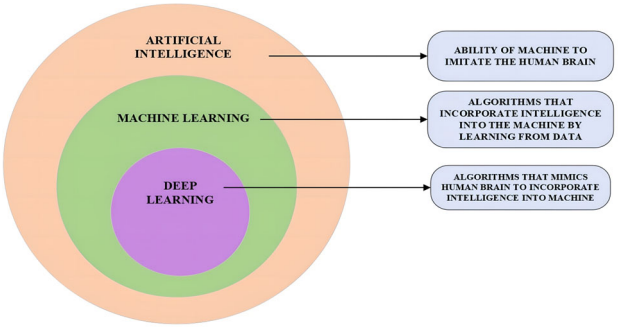


Fig. 1. Relationship between Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL).

Deep learning and machine learning are subsets of artificial intelligence. The Machine Learning (ML) algorithms are the subdivision of AI which have brought

Manuscript received February 22, 2024; revised May 20, 2024; accepted June 14, 2024; published January 17, 2025.

great changes in the healthcare industry, providing self-learning neural networks which can increase the quality of prediction by learning the external data features of a patient's medical screening images taken such as their Computed Tomography (CT) scan images, Ultrasound scan images, Magnetic Resonance Imaging (MRI) images and X ray images. ML algorithms are inbuilt with artificial neuron layers linked with one another which help in image identification, detection, classification, and segmentation. ML algorithms learn from a dataset of medical images which contain both normal and abnormal cell images and tries to predict the anomalies present in a quicker manner with accurate precision. Machine learning algorithms have played a major role in the field of oncology, pathology, and rare diseases.

A. Statistical Analysis on Health Care

Globally, the healthcare industry is currently undergoing a digital transformation. There seems to be no end in sight to the explosion of data in every aspect of our lives. Data from medical imaging tests, such as Computed Tomography (CT), MRI, and X-rays, significantly increases the volume of data. In the United States, CT scans rose from 3 million in 1980 to over 80 million by 2015. An average hospital produces 50 petabytes of data every year, of which a large amount comes from medical imaging. Digital pathology slides and X-rays are examples of 2D images that are frequently utilized in disease diagnosis. The creation of AI systems that can help radiologists and pathologists identify anomalies with high accuracy is made possible by the availability of massive datasets. AI-powered automated screening systems can swiftly evaluate thousands of 2D pictures, lessening the workload for medical personnel and improving early detection rates. 3D imaging techniques, such as functional MRI (fMRI), enable the study of dynamic processes in the body, such as brain activity. This leads to better understanding and treatment of neurological conditions. The study of dynamic bodily processes, such as brain activity, is made possible by 3D imaging methods like functional MRI (fMRI). This improves our knowledge of and ability to treat neurological disorders.

Large 3D datasets were used to construct AI algorithms that can identify minute differences in 3D scans that could go unnoticed in 2D pictures, increasing the precision of diagnosis.

The massive amount of health care data can be analysed with these Machine Learning (ML) algorithms to uncover hidden patterns and other useful information that traditional analytics are unable to find in a reasonable length of time. When the cases become more trivial and the accurate prediction of damaged cell pictures may sometimes involve a more complex but a better and good learning models involving deep learning models, that can interpret medical images like X-ray, MRI scan, CT scan and Ultrasound images.

B. Types of Medical Imaging

Medical Imaging is known as radiology where the doctors who are specialized in that are called radiologists. Medical practitioners use multiple imaging modalities to determine the abnormalities if any in our body parts. Multiple imaging methods are listed below:

- Ultrasound Imaging
- Computed Tomography (CT) imaging
- X-rays
- Magnetic Resonance Imaging (MRI)
- Positron–Emission Tomography (PET)
- Nuclear medicine imaging

Plain X-rays expose the structure and appearance of tumors, bones, and other dense matter. This is a simple, faster, and non-invasive method of imaging, there is a very minimal risk for the patients from exposure to ionizing radiation (X-rays) and the slight increased risk in for children. The CT scan produces cross-sectional layers with fine comprehensive images of the body parts like bones, organs, tissues, and tumors. Compared to x-rays, CT scans can detect wider range of anomalies with higher prediction rate by inducing higher radiation doses than the standard X-rays. In nuclear medicine imaging an injection, breathing, or ingesting a radioactive "tracer" is given such as Positron-Emission Tomography (PET). PET imaging is painless and can be used to diagnose, treat, or predict the result for a variety of disorders with the gamma-rays emitted to produce the images of bones and internal organs.

Ultrasound imaging (sonography) employs high-frequency sound waves to expose the internal organs within the body. These images are recorded in live environment so there is a possibility that they can capture blood flowing also through the body. There is no ionizing radiation exposure associated with ultrasound imaging and the physician can evaluate, view, diagnose and treat the medical conditions effectively without causing any harm to the patients. This is employed in cases involving breast examination, abdominal scanning, gynaecological, urological, and cerebrovascular examination. This medical imaging technique has been put to practice from 1980's and has given excellent safety record with its non-ionizing radiation. Since it has many advantages and safety parameters, we can use it for wide range of applications.

C. Ultrasound Image Processing

Ultrasound images are generated by sweeping the ultrasonic beam in the transducer. The transducer is rotated, tilted, and swinged mechanically across the body and a digital two-dimensional image of the body part is created using the signals that are processed further after being received from the transducer. Three dimensional images are further created by processing a series of adjacent two-dimensional images. Three-dimensional images live images of the beating of the heart and the fetus in the womb can be created. There are four modes in which ultrasound images are obtained and they are listed below:

- **A-mode Scanning:** A-mode scanning uses a single transducer to scan through the body and plots the echoes on the screen and represents the image in a one–dimensional representation format. It can identify specific tumors by employing the wave energy to focus with good accuracy [2].
- **B-mode Scanning**: The B-mode scanning is known as Brightness mode in which the organs and

tissues of the body parts are represented as points of varied brightness in the brightness mode, which uses a linear array of transducers to scan a plane through the body. As the image gets brighter it requires that more intense and focused is the reverberation of the sound waves [2].

- **M-mode Scanning**: In M-mode scanning called as motion scanning uses rapid sequence of B-mode images were the images sequentially follow each other on screen along a chosen ultrasound line, providing an monodimensional study of the heart to visualize and monitor the range of motion as the reflectors along the line are depicted on the time axis by the radiologists [2].
- **Doppler mode Scanning:** Doppler sonography generates images of blood flowing through blood arteries using high frequency sound waves. Doppler sonography can be applied to calculate the frequency shift for a specific jet of blood flow across a heart valve, and to determine and depict the jet's speed and direction [2].

II. Artificial Intelligence in Ultrasound Imaging

Over the past decade, artificial intelligence has thrown its light on medical imaging modalities, which has its roots in deep learning-based computer vision techniques. In the computer vision domain, the computers are trained to interpret and learn the digital images for object identification and classification problem. The dominant research areas in ultrasound imaging with AI assisted can be noise removal in ultrasound images, automated classification for breast, prostrate, liver, uterus and heart, segmentation of the abnormal area in the image. Ultrasound images, obtained from gynecological ultrasound, automatic detection of standard planes and the nature of the fetal and its anomalies can be identified. In radiology, AI has been involved a decade, but the dominance of AI has increased exponentially and it has deep impact on the accuracy of diagnosis with increased time efficiency [3].

A. Machine Learning in Medical Imaging

Machine learning has improvised the decision-making process in health care domain for the physicians by processing large datasets that lies beyond the scope of human capability and converts the analysis of that data made into clinical insights that aids the physicians in planning the treatment for the patients at lower cost and better ac-curacy. Machine learning algorithms provide an unbiased result and hence the efficiency and effectiveness are satisfied. To create an efficient ML algorithm, machine learning requires a specific amount of data, and feature engineering is required to extract the features. Feature engineering in machine learning relies on domain knowledge to identify relevant features and transformations. Without a good under-standing of the problem domain, it can be challenging to engineer effective features and may have to risk of over fitting. Feature engineering might introduce missing values or create features that are not available in the test set, making it challenging to deploy the model in real-world scenarios. Despite these drawbacks, feature engineering may remain an essential part of the machine learning workflow and it can significantly improve the model's performance and generalization capabilities. As the field of machine learning evolves, automated methods and techniques like deep learning may reduce the need for manual feature engineering in real life scenarios. The explosion of AI is being fueled by machine learning, which has been extensively used in radiology. Machine learning algorithms are now being superseded by deep learning models and algorithms so that more complex radiological tasks can be accomplished.

B. Deep Learning in Medical Imaging

Deep Learning algorithms have become inevitably popular in recent trend because of the abundant growth in the data in the big data field. The deep learning computing paradigm has become the most outstanding and widely used computational approach compared to ML and has achieved significant results in the field of computer vision outperforming the human intelligence [4]. Deep learning models have thousands of hidden characteristics that the modern graphics processing units and cloud architecture can support. Deep learning is being used in radiology to reveal clinically important features in imaging data, beyond what the human eye can see. The combination of deep learning with radiology can be popularly used oncology-oriented image analysis, due to its incredible diagnostic accuracy compared to earlier automated image analysis techniques known as Computer-Aided Detection (CAD) systems [5].

Deep learning is being applied in multiple domains which involves the processing, analysis of a vast set of data. The two key factors that are needed for DL techniques are the large amount of data and the large computational power. Deep Learning is frequently used in the healthcare industry to evaluate large volumes of data to aid with the early diagnosis of diseases and reduce manual processing [6]. Fig. 2 shows the health care applications that use the Deep Learning algorithms and the input data that must be given to the algorithms.

III. Convolutional Neural Networks in Ultrasound Imaging

Ultrasound medical images, which are non-invasive and harmless to human body and cost-effective are affected by many noises and the image quality is affected by the blood flow, position of the transducer. These factors tamper the quality of the image and corrupt it. So, in recent years, CNNs has been exploited in many ways to identify the abnormalities, classify it or to do segmentation of a particular Region of Interest (ROI). In paper [7], the Visual Geometry Group (VGG)-16T model was based on the VGG-16 architecture, to classify the thyroid nodules. Along with the fully connected levels, dropout layers and Batch Normalization (BN) layers have also been included. The performance of the VGG-16T model was assessed using a data set of grayscale US images from five distinct US machines.

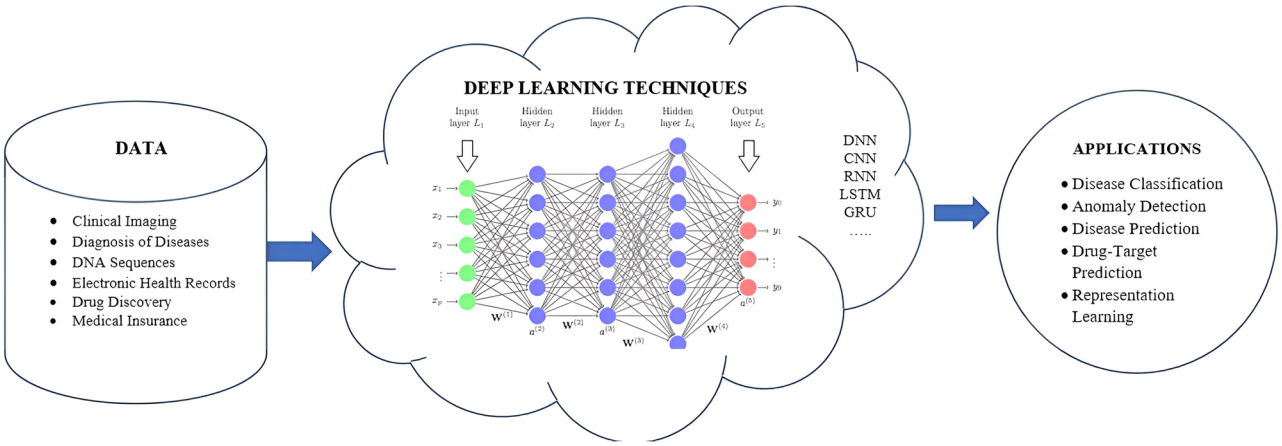


Fig. 2. Deep learning in health care applications.

The model performance was analysed using a data set of gray-scale Ultrasound images with 10-fold cross-validation. For this dataset, the VGG-16T model offered 87.43% sensitivity, 85.43% specificity, and 86.43% accuracy. This DL model helped radiologists identify thyroid cancer because it was precise and selective in differentiating between malignant and benign tumours [7].

In [8], a multi-task CNN model was constructed with 5911 ultrasound image datasets obtained from 2131 patients has served as the foundation for the CNN model. Hierarchical Loss (HL) function has been developed for classification and ALNM tasks. This CNN model had given a sensitivity of 83.5%, specificity 71.6% and Area Under the ROC Curve (AUC) of 0.878, respectively.

In [9], a comparative study on different pretrained CNNs classification models such as ResNet18, ResNet50, ResNet101, InceptionV3, InceptionResNetV2, GoogleNet, MobilenetV2, Squeeze Net, DenseNet201, and Xception. A dataset of 780 Breast Ultrasound images (133 normal, 437 benign, and 210 cancerous) was used for training and validation. The dataset's 375 Breast Ultrasound images were chosen at random to assess each CNN model's post-TL classification accuracy. Transfer Learning (TL) has been used to classify Breast Ultrasound images using the pretrained models.

A. Example for Feature Engineering with CNN

The authors Badža *et al.*, have worked on a labelled dataset containing MRI images of brain with different types of tumors such as glioma, meningioma, pituitary tumors were taken for classification. The machine learning models such as random forests, and Support Vector Machines (SVM) were used for extracting the features. To maximize the performance of machine learning models, the hyperparameters were adjusted using methods like grid search and random search. 10-fold cross-validation was used to assess the model's performance on folds of the dataset once the dataset has been divided into training and validation sets. The hyperparameters of the machine learning models are adjusted using methods like grid search and random search. With this specification the machine learning model achieved high levels of sensitivity, specificity, and accuracy. By doing feature engineering the model delivers state-of-the-art performance by utilizing deep learning representations in conjunction with a combination of texture-based, intensity-based, shape-based, and spatial data. This example shows that feature engineering has a significant impact on medical imaging and helps to improve clinical decision-making in neuro-oncology and patient care [10] .

In [11], the authors Usman *et al.*, have used a labelled dataset of retinal fundus images from diabetic patients were collected from ophthalmic clinics and hospitals. Statistical methods such as feature ranking and feature selection has been implemented by eliminating noise and redundant data. Principal Component Analysis (PCA) was used for dimensionality reduction of the feature space. The machine models were used for classification after hyperparameter tuning. This example portrays the importance of feature engineering with retinal fundus images for diabetic retinopathy.

Thus, we can infer that CNNs have been exploited in classifying and segmenting the ultrasound images. With this revolution of CNNs medical image classification and segmentation have rapidly developed in sensitivity and accuracy in classification. In our paper we have obtained 3D images in Neuroimaging Informatics Technology Initiative (NifTi-nii format) and processed using 3D CNNs.

B. Two-Dimensional Ultrasound Images

Two-dimensional Ultrasound imaging technology practiced worldwide as it is the safest imaging modality. These images are the most known format of flat images. This type of images is popularly used in monitoring the health of the fetus during the pregnancy phase of the women. Also, it is used in monitoring and identifying the anomalies of other various parts of the body. This 2D scans are the traditional imaging technique allow us to observe the interior organs, tissues, and blood vessels in their early stages and find aberrant structures like tumours, provide the most important information on the condition of the component under study. The computer records and analyses the high-frequency sound waves, producing 2D flat visuals of the graphs and images. The ultrasound images produced by computers are becoming clearer and of higher quality due to the exploitation of machine and deep learning algorithms. Research is still being done to further increase and improve the image quality, to get rid of the undesirable noise, and the processing speed.

C. Two-Dimensional Ultrasound Image Representation

The 2-dimensional images of the ultrasound scans are being retrieved in B-mode scanning, which is made up of vivid dots that are represented for the ultrasound echoes and are used for identification of lesions, cysts or tumors based on the structural anomalies. The images can have Tagged Image File Format (TIFF) or Portable Network Graphics (PNG) format or Graphics Interchange Format (GIF) format or Digital Imaging and Communications in Medicine (DICOM) format. The 2-dimensional images are present as raster images and the intensity of the pixel is dependent on the capability of the system. The resolution of the images is very low so we need to apply preprocessing techniques to increase the resolution of the image and these representations are considered while applying the deep learning algorithms. Raster images are made up of columns and rows of pixels. Each pixel depicts a regional characteristic, and its value reflects the pixel's conduct in that region. In medical images the grey scale image is popularly generated so the pixel values can be 0s or 1s. If the pixels in a grayscale image are of range from 0 to 255, then 0 denotes black and 255 denotes white. The raster images can have TIFF, PNG, GIFF or DICOM format. The structure of raster image is depicted in Fig. 3. The pixel representation in raster format for a 2-dimensional image is shown in the figure.

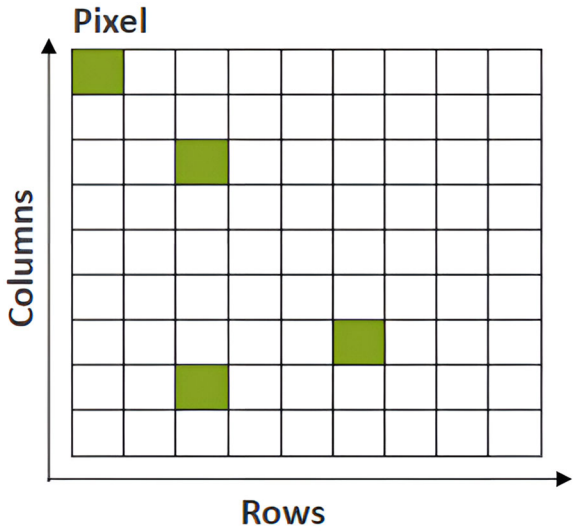


Fig. 3. Structure of raster image.

Medical images are of raster format and are represented in Binary Space Partitioning (BSP) model. The medical images are grey scale images and has black and white colors. So, the binary values 0 and 1 correspond to the appropriate colors black and white. When describing an image using a BSP representation in the shape of a tree, Tf, node of the tree is called as a "container" for data in each of the image's partitioned regions. The first partitioning line and the average colour of the entire image are both located in the root node. Only the region's colour information is contained in the nodes that correspond to non-partitioned (cell) areas [12]. Fig. 4 depicts the tree structure Binary Space Partitioning model of a 3D image.

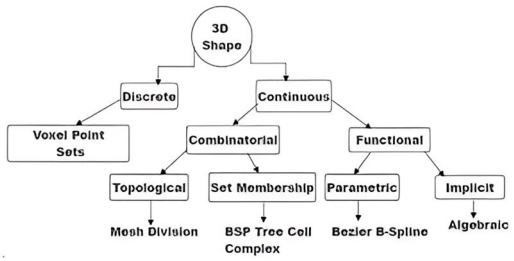


Fig. 4. Binary Space Partitioning (BSP) representation of an image.

The image can be represented as a 2D array or matrix. For the grayscale image "*I*" "*I(x, y)*" signifies the pixel value at coordinates *(x, y)* in the image. For a grayscale image representation, the equation is given as:

$$I\,(x,y) = Grayscale\ value\ at\ pixel(x,y) \quad (1)$$

The equation for a 2D grayscale image representation is given as:

$$I(x,y) = f(x,y) \quad (2)$$

In the Eq. (2), *f(x, y)* is the function that converts the coordinates *(x, y)* to the pixel intensity value, and *I(x, y)* is the pixel intensity at coordinates *(x, y)* in the image. This equation simply states that the grayscale image "I" is represented by its pixel values at each coordinate *(x, y)* in the 2D space.

D. Three-Dimensional Ultrasound Images

Early applications of 3D ultrasound technology date back to the late 1970s, and they have expanded ever since. It first had a significant diagnostic influence in obstetrics, and it is now starting to demonstrate its value in gynaecology as well. While 3D ultrasound applications in gynaecology are still in their infancy, we anticipate that they will improve medical ultrasound's diagnostic imaging [12]. The influence of technology in medicine is larger than ever and continues to grow because to improvements in processing power and algorithms. The use of three-dimensional US imaging in medical diagnosis has become worldwide as it is minimally invasive and image-guided procedures and has demonstrated to become more and more significant. The 2D ultrasound images in a 3D ultrasound examination to provide an objective 3D depiction of the anatomy and disease. The doctor can then use the same or a different computer to view, edit, and measure this image in 3D. Additionally, a 2D cross-sectional image can be created at any anatomical point in any orientation, without restrictions, and it can be simply registered with a preceding or following 3D image [13].

E. Three-Dimensional Ultrasound Image Representation

Three-Dimensional image representation involves encoding and displaying visual information in a three-dimensional space. In a variety of industries, such as computer graphics, virtual reality, medical imaging, and 3D modelling, it is frequently used to represent objects, scenes, and settings. The terms width, height, and depth

are used to describe the three spatial dimensions. Everything we observe in the physical world is three dimensional. Generally, 3D images are represented in the form of voxel grids. Voxel grids are 3D analogy of 2D pixel grids. They represent a 3D volume by dividing it into a regular grid of small cube-shaped elements called "voxels." Each voxel contains information such as color, density, or material properties, and they collectively form a volumetric representation of the 3D scene. Volume rendering is a technique for visualizing the voxel, volumetric data, such as medical CT or MRI scans. It involves projecting the 3D data onto a 2D image plane, considering properties like density and opacity to create a 3D-like representation. The taxonomy of 3D representation is shown in Fig. 5 below.

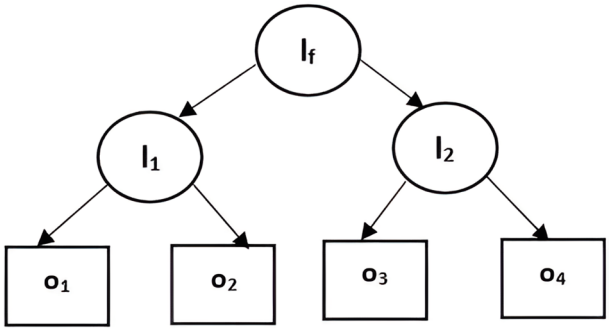


Fig. 5. Taxonomy of 3D image representation.

Generally, 3D medical images are Discrete and hence we can represent them using voxel point sets. Volume rendering and volumetric segmentation, are powerful tools is used for visualizing 3D medical data and to identify and isolate specific structures or regions of interest. Volume rendering allows clinicians to observe the internal structures of organs and tissues in a 3D context. Segmentation is critical for quantitative analysis and surgical planning. By combining multiple 2D slices from different imaging modalities, 3D reconstructions can be generated, providing a comprehensive view of complex structures or tumors. Fig. 6 depicts the 3-dimensional view of a 3D image along its axis x, y and z axis.

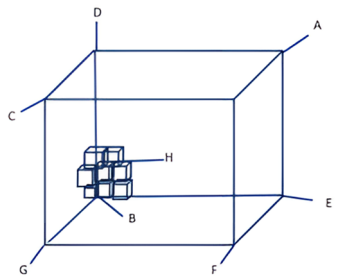


Fig. 6. Model diagram of a 3D image.

The following Fig. 7 shows the 3-dimesional view of an Ultrasound image. The slicer tool generates the 3D view of the Ultrasound Image. Then the 3D images were convertered to 2D slices. The 2D slices along the x-axis,y-axis and z-axis are depicted in Fig. 8(a)—x-axis , Fig. 8(b)—y-axis and Fig. 8(c)—z-axis.

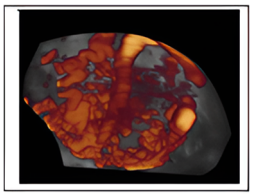


Fig. 7. 3D view of an ultrasound image.

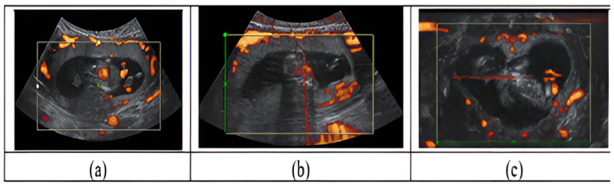


Fig. 8. Two dimensional slices of 3D images along the (a) x-axis (b) y axis (c) z-axis.

Medical imaging techniques like Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) frequently employ voxel-based representations. Assuming we have a 3D medical image volume, and each voxel at coordinates (*x, y, z*) contains an intensity value denoted as "*I (x, y, z)*".

The equation for voxel-based representation is expressed as:

$$Voxel\ Representation : I(x, y, z) \tag{3}$$

In the Eq. (3), *I (x, y, z)* signify the intensity value at the voxel in the 3D volume with coordinates *(x, y, z)*. In volumetric rendering, transfer functions and interpolation methods are used to convert the voxel intensities into color and opacity values for rendering the 3D volume.

F. Comparison of 2D/3D Medical Images

Though 2D Ultrasound are traditional scanning done by the medical practitioners to understand the human anatomical features, it is difficult to get the Region of Interest (ROI) [12]. Three-dimensional (3D) intraoperative US volumes are created from two-dimensional (2D) slices to improve surgical planning and increase accuracy in the coronal plane, which can occasionally be difficult or impossible to view on 2D images due to anatomical constraints [13]. For radiologists, patients, and referring doctors, 3D images are quicker, easier to read, and produce precise anatomical visualizations. They also result in lower expenses for the healthcare system. With the rapid advancement in Machine and Deep learning techniques of artificial intelligence, CNN's have exploited the medical image analysis for classification, segmentation, detection, and localization of medical images like CT, MRI, X-rays, and Ultrasounds. In this work, we have explored the features of 3D CNN's and compared the workings of CNN's on 2D medical ultrasound images and 3D medical ultrasound images.

IV. Methods and Techniques Used

The work has been implemented using Google Collaboratory. The dataset was loaded in google drive in 3D nii format images and the preprocessing of the images

were done. Then we have converted the 3D nii format images into 2D images in JPEG format using the Slicer 4.11 tool and these 2D images were loaded in a separate folder in the google drive and preprocessed and given as input a 2D CNN which is sequential CNN and the outputs were compared.

A. Dataset Description

3D volumes of ultrasound image data have been obtained in NifTi (Neuroimaging Informatics Technology Initiative) format from the John Radcliffe Hospital, University of Oxford, Oxford, OX3 9D. The 3D volumes of data have been pre-processed and annotated. The 3D images were also converted to 2D images using the Slicer tool and those 2D data has been processed and annotated. These images contain data about the uterus and the fetus during the first trimester of the pregnancy. As the first trimester is the most crucial time of pregnancy, if prediction can be done on any anomalies or irregularities, it would be highly beneficial in identifying the status of the mother and the baby for the doctors and gynaecologists. Fig. 3 represents the 3D view of the NifTi image of our dataset and Fig. 4 represents the 2D slices of the 3D image taken at 3 different axis such as x axis, y axis and z axis. The dataset was split into 70:30 ratio for training and testing. There were totally 640 3D images of which 448 were given for training and remaining 192 were used for testing. Cross Validation can be done using the k-fold cross validation as we have a small dataset.

B. Two Dimensional /Sequential CNN

3D volumetric data was converted to 2-Dimensional data in JPEG format using slicer tool. The 3D Volumes of data was cut in 3 angular positions along the x, y, and z axis. The raw 2-Dimensional data was then preprocessed and fed into a sequential 2-dimensional CNN. CNNs have contributed towards medical image processing for image classification, segmentation, and anomaly identification. In the Brain Tumour Segmentation (BRATS) process, CNN has been used. Ischemic Stroke Lesion Segmentation (ISLES), and Multimodal Brain Tumour Segmentation [14].

CNN has also established itself as an unavoidable option for comprehending medical images such as tumour detection, tumour categorization, and malignant vs. benign tumours [15] , the detection of optical coherence tomography picture [16] , the detection of anomalies of the heart [17], breast, chest, eye, and so forth have all been accomplished using CNNs in the field of medical image interpretation. Covid chest X-rays and CT images have been classified in large part thanks to CNNs. A 2D CNN processes 2D input data, such as grayscale or RGB images with convolutional layers to learn and extract relevant features automatically [18]. The key equations involved in a typical 2D CNN are described below:

1) Convolution operation

The convolution operation is the fundamental building block of a CNN. To create feature maps, a tiny filter known as the kernel, is slid across the input image while computations are made for element-wise multiplication and summing. Assume we have a filter "*W*" and an input picture "*X*" The following is a representation of the convolution operation for a single feature map "*F*" at point *(i, j)*.

$$F(i,j) = \sum\sum(X(x,y) \times W(i+x, j+y)) \quad (4)$$

F (i, j) signifies the feature map value at *(i, j)*, *X (x, y)* represents the input image pixel value at *(x, y)*, and *W (i+x, j+y)* is the filter weight at position *(i+x, j+y)* in the Eq. (4).

2) Pooling operation (max pooling)

Pooling layers minimize the computational complexity by extracting only the important information and lower the dimensions of feature maps in the spatial manner. Max pooling is the pooling operation that takes the maximum value from a small window of the feature map. The pooling operation at a specific position *(i, j)* is expressed in the Eq. (5).

$$Max_{pooling}(i,j) = \max\big(F(i',j')\big) \quad (5)$$

where, $i' \epsilon [iS, iS+P]$ and $j' \epsilon [jS, iS+P]$, the value at location *(i, j)* in the pooled feature map is Max_Pooling *(i, j)*, and the values in the feature map are *F (i', j')*. "*S*" is the stride which tells the number of windows shifts for each pooling operation, and *P* is the size of the pooling window.

3) Activation function

In order to induce non-linearity and capture complex patterns in the data, an activation function is applied element-wise after each convolution process. The Rectified Linear Unit (ReLU), which is used most frequently as an activation function, is expressed in equation as below:

$$ReLU(x) = \max(0, x) \quad (6)$$

4) Fully connected layer

The feature maps are flattened into a one-dimensional vector and given as input into a fully connected layer after many convolutional and pooling layers. The completely connected layer's equation expressed as below:

$$Y = W \times X + b \quad (7)$$

In the Eq. (7), *Y* represents the output vector of the fully connected layer, *X* represents the input vector of the flattened feature maps, *W* stands for the fully connected layer's weight matrix, and *b* for the layer's bias vector. These are the key equations involved in a basic 2D CNN architecture. The flow diagram of the sequential CNN is depicted in Fig. 9.

Numerous literatures are available on the efficiency and the widespread use of 2 dimensional CNNs. Structure of 2D CNNs is shown in Fig. 8. Here in our work, we have converted the 3D images to 2D slices and applied preprocessing techniques on it. The images were rescaled to 150×150 shape and were directly given to CNN network for classification.

The 2D CNN had 2 convolution filters of 64 × 64 and a max pooling layer of 64×64 and was normalized and flattened to 32×64 image size. The total trainable

parameters were 280,897 and non-trainable parameters were 128. The classifier accuracy was 55.65% and loss was 39.13%.

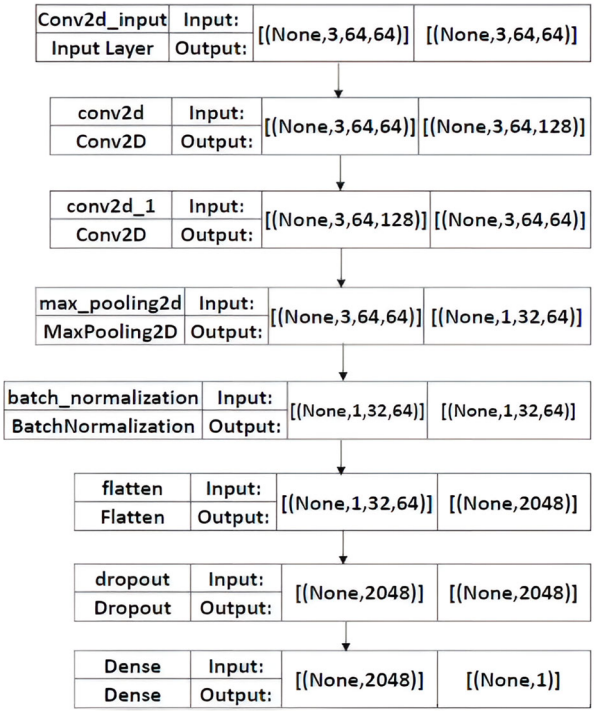


Fig. 9. Flow diagram structure of 2-dimensional CNN.

C. *Three Dimensional/Sequential CNN*

Three dimensional CNN are required for data analysis when we have volumetric and temporal data. This model as an automated tool for abnormality detection and evaluation in medical imaging. In a specific context, 3D CNN analyses a sequence of frames or images. While 3D CNNs are comparable to 2D CNNs, they may learn representations of volumetric data since they accept 3D volumes as input. It is highly beneficial for medical images as they are more complex and volumetric. The 3D CNN produces a 3D activation map during the convolution step, which is required for time and volumetric context as well as data interpretation. For identifying the elements at a low level, a three-dimensional filter applied in the x, y and z directions with respect to the dataset's 3D convolution step [19]. A cubic shape or a cuboid-shaped 3-dimensional volume space is obtained as a output and hence, three dimensions are present in both the input and output. By down sampling an input representation in 3D space, the 3D max pooling layer decreases its dimensionality and makes assumptions about the features that are present in the sub-regions binned.

In this work we have used 3D-CNN in which the first 2 layers used 64 filters with 3D Kernel. The activation function employed here was ReLU (Rectified Linear Unit). The third layer has 128 filters and fourth layer,256 filters with same 3D Kernel and the activation function "ReLU". Batch normalization and a 3D Max Pooling Layer has been employed. The output layers are dense with "sigmoid" as the activation function. The 3D convolution operations and equations are described below. Fig. 10 portrays the overall architecture of a 3D Convolutional Neural Network. The recent works on 3D CNN has been elaborated in Table I. The papers show the different datasets used and the metrics used in evaluating the performance of 3D CNN.

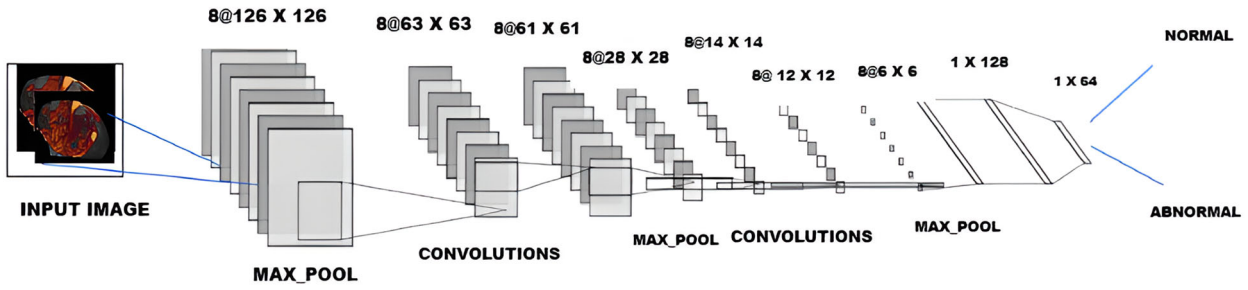


Fig. 10. Architecture diagram of 3D-convolution neural network.

TABLE I. RECENT RELATED WORKS OF 3D CNN

S. No	PAPERS			
	Deep Learning Method Adopted	**Description**	**Metrics Analysed**	**Merits And Demerits**
1	Adaptive 3D CNN [20](2024)	Diagnosis of Alzheimer's Disease		An ensemble model was developed with 3D CNN and Genetic algorithm was used to differentiate the different classes of Alzheimer's disease.
2	2DCNN, Inception V3 Model with Transfer Learning [21] (2024)	Monkey Pox Diagnosis	Accuracy: 2D CNN-86% Accuracy: Inception V3-98%	Can be globally used for image diagnosis Handles high volumes of images
3	Lite 3D CNN [22] (2023)	Human Movement Recognition	Accuracy: 3D CNN-91.6%	Optimizes the traditional three-dimensional convolution network

S. No	PAPERS			
	Deep Learning Method Adopted	**Description**	**Metrics Analysed**	**Merits And Demerits**
4	ResNet Based 3D CNN [23] (2023)	Action Recognition with fusion BERT	Accuracy: ResNet 3D CNN-82.53%	3D CNN are used to filter spatiotemporal features in the video.
5	3D CNN with batch input [24] (2021)	Classification of Coronavirus and Diagnosis	Accuracy: >99%	The covid affected CT images were super-imposed into many depths of volumes and were used to classify covid 19.
6	Shallow 3D CNN [25] (2020)	Brain Hemorrhage detection	Avg F1-Score-0.77	3D volumetric scans were normalized with intensity of training samples and the contrast was increased on the abnormal region of CT scans.
7	3D Deep Add Net (3D DANET) [26] (2018)	Brain Image Classification on multicenter datasets	Accuracy: 3D DANET - >92%	Can be globally used for Intelligent medical treatment and clinical practices.
8	3D CNN HadNet [27] (2018)	Alzheimer's Disease Diagnosis	Accuracy: 3D CNN HadNet-88.31%	The MRI data was processed as a whole and based on the learned features the data was classified.
9	CAD Based 3D CNN [28] (2017)	Lung Cancer Classification	Accuracy: 3D CNN-86.6%	The DSB dataset was segregated into minibatches of size 1,10,50 and 100. If the batch size was increased. The images were fed to these 3 layered CNN networks.
10	3D CNN with Conditional Random Fields [29] (2017)	Brain Lesion Segmentation	Accuracy: 3D CNN-83.6%	The model was specific to the dataset used. Was not able to handle heterogenous data.

1) 3D convolution operation

The 3D convolution operation is the primary building block of a 3D CNN. It works by applying a 3D filter (kernel) to the 3D image data and perform multiplication and summation to create feature maps for each element. On an assumption if we have an input 3D volume "*V*" and a 3D filter "*W*" The 3D convolution operation for a single feature map "*F*" along the coordinates *x*, *y*, *z* axis can be expressed in Eq. (8) as:

$$F(x,y,z) = \sum\sum\sum\big(V\,(i+x, j+y, k+z) \times W(i,j,k)\big) \quad (8)$$

In the Eq. (8), *F (x, y, z)* is the feature map at position *(x, y, z)*, *V (i+x, j+y, k+z)* signifies the voxel value in the input 3D volume at position *(i+x, j+y, k+z)*, and *W (i, j, k)* is the weight in the 3D filter at position *(i, j, k)*.

2) Activation function

3D CNN uses the same 2D CNNs, an activation function is applied element-wise after each 3D convolution to introduce non-linearity. Rectified Linear Unit (ReLU) was used as an activation function to extract the complex patterns in the three-dimensional data.

$$ReLU(x) = \max(0, x) \quad (9)$$

3) 3D pooling operation

Pooling layers in 3D CNNs lower the dimensions along the *x, y, z* axis in the spatial domain. These extract only the important information and thereby reducing the computational complexity. The pooling operation along the coordinates *x, y, z* axis can be expressed in Eq. (10).

$$Max_{Pooling}(x,y,z) = \max\big(F(i,j,k)\big) \quad (10)$$

were, $i \in [xS, xS+P], j \in [yS, yS+P], k \in [zS, zS+P]$. The Max_Pooling operation at *x*, *y*, *z* dimensions represent the position value of along the three axes in the pooled feature map, the function *F (i, j, k)* are the values in the feature map, "*S*" is the stride, and P is the pooling size of the window in all three dimensions.

4) 3D fully connected layer

The feature maps are flattened into a 1D vector and fed into a fully linked layer after many 3D convolution and pooling layers. The following equation represent the 3D fully connected layer.

$$Y = W \cdot X + b \quad (11)$$

In the Eq. (11), *W* is the weight matrix of the fully connected layer, b is the bias vector of the fully connected layer, Y is the output vector of the fully connected layer, *X* is the input vector on flattened feature maps).

5) Three dimensional CNN-related works

Chen *et al.* [30] in their paper have used 3-dimensional convolution neural networks for brain tumor segmentation using the brain MRI volume images. To achieve correlation between adjacent slices the 3D U-Net and V–Net had been applied and the three-dimensional convolutions are obtained as results. 3D networks suffered from high computational overheads due to multiple 3D convolutions. To bridge the gap between the 3D and 2D convolutions they a dilated three-dimensional Multi-Fiber Network (DMF Net) was proposed in paper [31], which was developed on top of a multi-fiber unit. This network had a group convolution framework, that had weighted three-dimensional dilated convolution operation to gain multi-scale image representation for segmentation. Lee *et al.* [32] in their research work have automatically

segmented the pulmonary lobes of the lungs from the CT images using a high-resolution and effective 3D fully convolutional network. The Pulmonary Lobe Segmentation Network (PLS-Net) was the name given to the network, which was created with the goal of effectively using 3D spatial and contextual information from high-resolution volumetric CT scans for volume-to-volume learning and inference. In [33] Yang et al has done a multi-institutional dataset of 210 CT pictures from individuals who had a variety of lung problems was used to test the PLS-Net. The results of the experiments demonstrated that our PLS-Net achieves cutting-edge performance with greater computing efficiency. Kruthika *et al*. [34] used a preprocessed MRI dataset of the ADNI dataset and an auto-encoder to produce the input features of 3D patches. Compared to 2D CNNs, the classification accuracy was higher at 89.1%. Feng *et al*. [35] employed stacked RNN layers on top of a 3D CNN for the categorization of Alzheimer's disease using PET and MRI data. The accuracy of their model, which used 3D CNN and LSTM, was 65.5%. Wegmayr *et al.* [36] used a deep 3D CNN on ADNI and AIBL dataset of nearly 20000 T1 scan images and obtained a accuracy of 72% for mild cognitive impairment, 86% was classified as Alzheimer's disease and have reached a classification accuracy of 94.85% using a 3D CNN to predict the start of Alzheimer's disease even in a single participant using resting state fMRI data. Nie *et al.* [37] has applied 3D CNN with supervised features on the brain tumor images which consists of T1 MRI, fMRI and DTI formats of images and have obtained a classification accuracy on 89.85%. Amidi *et al.* [38] has used a 2-layer 3D CNN on the Protein shape dataset, which contains 63,558 enzymes from PDB databases, and achieved a classification accuracy of 78%. Thus, 3D CNNs has been exploited and applied to the classification of medical images. Medical image segmentation, detection, registration, and localization have also benefited from its expansion [38, 39]. Though it has high computational overheads, it highly useful in extracting the sensitive medical features and to improve the accuracy of classification of medical images.

In our work, we have preprocessed our dataset and have given the data in a 3D CNN for binary classification.

D. 3D CNN-Preprocessing Techniques Applied

1) Image normalization

This is the primary step applied on our dataset. On completion of this task, we can get two effects: it increases the contrast between the bright and dark structural components that make up the texture under investigation, and it lessens the influence of the local mean intensity. Both impacts typically raise the calibre of the characteristics acquired. Our image dataset when normalized the pixel values of the images were standardized. The standard deviation of pixel values of all the images were calculated for the black and white color channel and each pixel value is divided by the standard deviation of its color channel. Normalization ensures that the pixel values are consistent and have a uniform distribution which helps the CNN models have a better performance.

2) Image resizing

Image resizing applied to modify the total number of pixels in an image. Our dataset of volumetric images was rescaled along the z axis and are resized as 64×128×128-pixel images. The resizing of our dataset ensured that the aspect ratio of the image are preserved and there was no distortion.

3) Gaussian filtering

Gaussian filters blur the images and remove noise from the images. A single dimensional function for gaussian filter is expressed in Eq. (12).

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{x^2}{2\sigma^2}} \tag{12}$$

When gaussian filters are applied on images a 2-dimensional gaussian function has been used. This is a one dimensional, two gaussian function's product, and is given by:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{13}$$

Gaussian filters are non-uniform low pass filters in which the kernel coefficients get smaller from the centre. The weighting of central pixels is greater than that of peripheral pixels. Wider peaks (more blurring) are produced by bigger values of. The kernel size of the Gaussian filter keeps increasing in size for getting the smoother effects. Controlling th kernel size and applying sigma parameter has smoothened the edges and prevented distortion. We have provided 448 volumes to training after applying the preprocessing procedures, and 192 volumes to validation. On applying the gaussian filter on our dataset reduced the noise in the images and preserved the edges The output of the preprocessing steps is given in Fig. 11.

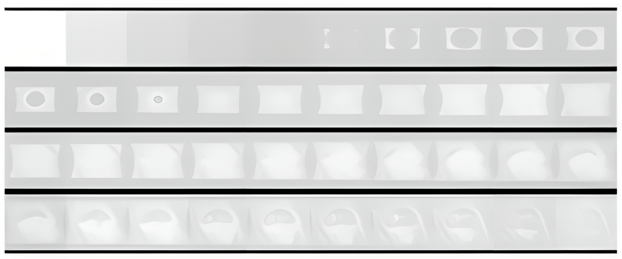


Fig. 11. Pre-processed 3D data.

E. Experimental Setup

The 3D CNN model framework was adopted for our experimantal analysis. The 3D CNN processes 3 dimensional volumetric images by learning the hierarchial representations of the spatial and tempoaral features. The basic framework was used for our analysis in which the 3D images after preprocessing has been given as input and the accuracy was compared with the normal 2D CNN. Without gaussian filtering when the data was given as input to a 3D CNN the classification accuracy was 0.7766, when the 3D CNN was run for 50 epochs. Then we applied gaussian filtering and then when the pre-processed data was given as an input to a 3D CNN which has total params: 1,352,897

in which the trainable parameters are 1,351,873 and non-trainable parameters are 1,024.

1) *Model tuning*

The existing 3D CNN was enhanced by adding convolution layers for performing the complex medical image classification. The architecture of the model was enahnced by adding convolution layers and executed for 50 epochs in which the classification accuracy obtained was 0.7979 for training and the validation accuracy was 0.6452. The graph obtained is as shown in Fig. 7 for accuracy and loss. The base layers were also trained with a lower learning rate. This fine tuning helps in learning the feature representations for the complex 3D medical images.

2) Validation and generalization

Model validation and generalization are essential components that guarantee the model's dependability and efficiency for medical imaging applications. The dataset that was divided into training and validation sets. The dataset has been divided into five folds and then validated. For each iteration one-fold had been used for training and remaining were used for validation. The model had been validated with the validation data and the metrics such as accuracy, precision and have been evaluated. Generalization refers to the ability of the model to perform well on the unseen testing data. Mathematically, generalization of a model can be evaluated and analysed with the following metrics. The minimization of loss function $\mathcal{L}$ added with a regularization term to avoid over fitting is given as:

$$\mathcal{L}_{Total} = \mathcal{L}(\hat{y}, y) + \lambda.\mathcal{R}(W) \tag{14}$$

here, $\hat{y}$ was the predicted output and y is the actual output. $\mathcal{L}$ is the cross-entropy loss function and $\mathcal{R}(W)$ is the regularization term that was used. The cross validation done using 5-fold validation score which estimated the generalization performance:

$$CV\ Score = \frac{1}{k}\sum_{i=1}^{k}\mathcal{L}(\hat{y}_i, y_i) \tag{15}$$

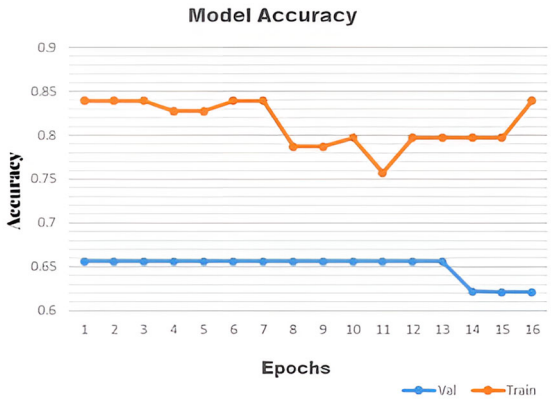
here, k represents the number of folds and for our experiments we had used 5 folds. $\hat{y}_i\ and\ y_i$ are the predicted and actual outputs. This score helps to understand the variance and bias of our 3D CNN model. The generalization model error can be decomposed into bias and variance as:

$$Gernalization\ Error = Bias^2 + Variance + Noise \tag{16}$$

Bias is the error introduced from the dataset to the model which can increase the complexity of the model. Variance is behaviour of the model for different datasets and the noise is the unwanted signals present in the dataset already. Thus, generalization of our model is based on the above components.

The accuracy obtained on using our model has been represented in the Figs. 12(a) and 12(b).

(a)



(b)

Fig. 12. Output Graph of 3D CNN. (a)3D CNN Model Accuracy Graph, (b) 3D CNN Model Loss Graph.

The accuracy obtained from our model 3D CNN is 77.79% without applying gaussian filter mechanism. After applying preprocessing techniques like normalization, resizing and gaussian filtering for noise removal the accuracy was increased to 79.78%.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{17}$$

here, TP-True Positives, TN-True Negatives, FP-False Positive, and FN-False Negatives which were classified by the 3D CNN classifier. The metrics used for evaluating the performance of our model are precision, recall and F1-Score.

$$Precision = \frac{TP}{TP+FP} \tag{18}$$

Precision value determines the actual positive cases that the model was able to predict. This value determines whether the model can be reliable or not. Recall captures all the actual positive cases and it must avoid predicting the abnormal cases as normal.

$$Recall = \frac{TP}{TP+FN} \tag{19}$$

The Table II below shows the Precision and Recall values of the 3D CNN model after preprocessing is shown below.

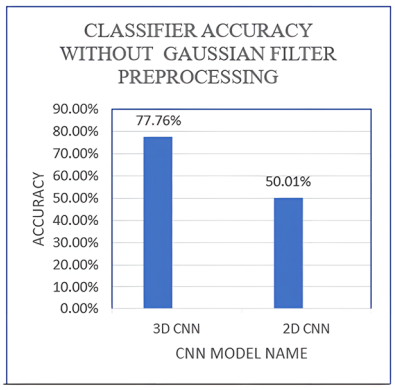
TABLE II. EVALUATION METRICS OF 3D CNN

Class name	Metrics		
	Precision	Recall	F1-Score
Class 0	0.7957	0.8031	0.7994
Class 1	0.8000	0.7926	0.7963
Accuracy		0.7978	
Misclassification Rate		0.2022	

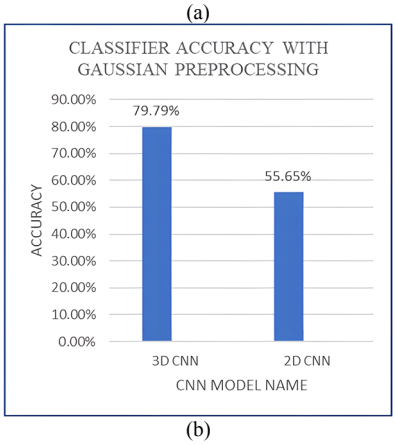
Table III shows the confusion matrix of our model. Our model performed well with respect to metrics like precision, recall and F1-Score. These metrics used in our evaluations provide a comprehensive review of the strength of our model. With preprocessing and performing hyperparameter fine tuning of our model the accuracy of the model is being increased. The model supports that when the medical images such as ultrasound, MRI or CT images when given as 3D images as input will be able to perform classification with improvised accuracy rather than giving 2D images as input and performing classification with linear CNN (2D CNN).

TABLE III CONFUSION MATRIX OF OUR MODEL

Class Name	Predicted	
	Normal	Abnormal
Normal	0.4028	0.1034
Abnormal	0.0988	0.3951



(a)



(b)

Fig. 13. Bar chart representation of accuracy metric of 3D CNN and 2D CNN. (a) Accuracy Comparison Graph for 2D/3D CNN without gaussian filter. (b) Accuracy Comparison Graph for 2D/3D CNN with Gaussian Filter.

The Fig. 13 represents the bar chart comparison of the accuracy metric obtained in both 3D CNN and 2D CNN. It can be observed that after preprocessing with gaussian filter 3D CNN had given higher accuracy in classification than the 2D CNN. Also, after preprocessing the classification accuracy has improvised.

V. COMPARISON OF RESULTS

From our obtained results we can find that 3D CNN had given better accuracy in classification of 3D ultrasound images than the sequential 2D CNNs. Thus, we can see that in medical imaging of ultrasound images can be directly given to 3-dimesional CNN for binary classification to have better performance. A 3D convolution operation's time complexity is influenced by the size of the input volume, the size of the convolutional kernel, and the quantity of filters. The time complexity of a single 3D convolution operation is expressed as $O(C \times D^3 \times H^3 \times W^3)$, where D, H, and W stand for the depth, height, and breadth of the input volume, respectively, and C is the quantity of input channels. When implementing our 3D CNN model, the time taken for running the epochs was higher than when it was executed on sequential CNN. 3D CNNs have a higher time complexity than 2D CNNs due to the additional dimension involved in the convolutional and pooling operations.

1) Advantages of 3D CNN and 3D diagnostics

The increase in computational complexity can make training and inference in 3D CNNs more computationally demanding, requiring more resources such as GPUs for efficient processing. However, 3D CNNs are essential for tasks that involve volumetric data, such as medical imaging or 3D video analysis, where they can capture spatial relationships and patterns in three-dimensional space. For 2D image data, 2D CNNs are still widely used and remain computationally efficient for image recognition and related tasks. The translation invariance of 3D CNNs' convolutional operations allows the model to identify patterns in the input data regardless of where they are located spatially and captures spatial relationships between the voxels in three dimensions.

With so many advantages for healthcare practitioners, 3D imaging has completely changed the medical field. Treatment planning, patient outcomes, and diagnostic accuracy are all improved by 3D imaging, which offers a more thorough image of anatomical structures. To schedule surgery and administer radiation therapy for cancer treatment, for example, 3D CT and MRI images are crucial. AI algorithms developed using large 3D datasets can detect subtle changes in 3D scans that might be missed in 2D images, improving diagnostic accuracy.

Healthcare is changing because of the growing amount of 3D medical picture databases, which make it possible to conduct more sophisticated research and provide more individualized treatment plans and precise diagnosis. The potential to enhance patient outcomes and increase the effectiveness of healthcare delivery will grow as these databases continue to develop.

2) Comparison of 3D CNN / 2D CNN

3D CNN slides over 3D volumes and captures both spatial and temporal information and also can dynamically adopt to the changes over time whereas 2D CNN slides and captures only the spatial data and have poor generalizability over the dynamic or time series data. 3D CNN are deeper than 2D CNN are shallow because the 3D CNN handles additional dimension. The total number parameters in 3D CNN are very much higher than 2D CNN. The 3D resampling and augumentation have been done as preprocessing in 3D CNN for handling the data efficiently. The memory requirement and the training time of 3D volumetric images with the 3D CNN are very much higher due to the increased complexity in parameter space whereas the 2D CNN has smaller dimension and also may require lower memory requirement.

VI. Conclusion

In this paper we have reviewed the characteristics of ultrasound images, 2D CNN and 3D CNN and have described the features of our dataset. The Experimental results were obtained by executing our dataset with 2D and 3D Convolutional Networks. From the results it was inferred that the for the medical ultrasound images 3D CNN is well suited for performing binary classification. Though the time complexity of 3D CNN is higher, we can prefer 3D CNN as its accuracy is higher and highly beneficial in 3D volumes medical ultrasound image classification when compared to 2DCNN. The findings suggest that 3D CNNs are particularly advantageous when volumetric data is available, delivering higher accuracy and robust feature representations. However, researchers and practitioners should be mindful of the increased computational demands of 3D CNNs but these computational overheads can be easily overcome by providing high end systems for analysis and diagnosis. Our model will be highly beneficial to the medical professionals for their diagnosis as the 3D images can be easily classified with better performance compared to 2dimensional images.

VII. Future Enhancements

The 3D CNN architectures can be refined and to improve their accuracy and efficiency in handling the medical images. The 3D CNN can be robust and can be generalized across various platforms. Future research can explore on increasing the robust features of 3D CNNs such as data augmentation and can be adopted for each domain specific applications. If deployed in real world, there will be a good impact on ultrasound imaging and other medical imaging modalities. In future research the time complexity of handling 3D images on the 3D CNN can be reduced and can be used in medical imaging real world applications.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Angelin Beulah S.: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation; Sivagami M.: Visualization, Investigation, Reviewing and Editing. All authors had approved the final version.

References

[1] A. Bohr and K. Memarzadeh, "The rise of artificial intelligence in healthcare applications," *Artificial Intelligence in Healthcare*, pp. 25–60, 2020. doi: 10.1016/B978-0-12-818438-7.00002-2

[2] A. Carovac, F. Smajlovic, and D. Junuzovic, "Application of ultrasound in medicine," *Acta. Informatica Medica.*, vol. 19, no. 3, p. 168, 2011. doi: 10.5455/aim.2011.19.168-171

[3] W. B. Shuo, L. M. Ji-Bin, Z. M. Ziyin, and E. P. John, "Artificial intelligence in ultrasound imaging: Current research and applications," *Advanced Ultrasound in Diagnosis And Therapy*, vol. 3, no. 3, p. 53, 2019. doi: 10.37015/AUDT.2019.190811

[4] L. Alzubaidi *et al.*, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021. doi: 10.1186/s40537-021-00444-8

[5] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthc. J.*, vol. 6, no. 2, pp. 94–98, Jun. 2019. doi: 10.7861/futurehosp.6-2-94

[6] S. Bhattacharya *et al.*, "Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey," *Sustain Cities. Soc.*, vol. 65, 102589, Feb. 2021. doi: 10.1016/j.scs.2020.102589

[7] Y.-C. Zhu, P.-F. Jin, J. Bao, Q. Jiang, and X. Wang, "Thyroid ultrasound image classification using a convolutional neural network," *Ann. Transl. Med.*, vol. 9, no. 20, pp. 1526–1526, Oct. 2021. doi: 10.21037/atm-21-4328

[8] C. Li *et al.*, "A convolutional neural network based on ultrasound images of primary breast masses: Prediction of lymph-node metastasis in collaboration with classification of benign and malignant tumors," *Front. Physiol.*, vol. 13, Jun. 2022. doi: 10.3389/fphys.2022.882648

[9] S. M. Badawy, A. E.-N. A. Mohamed, A. A. Hefnawy, H. E. Zidan, M. T. GadAllah, and G. M. El-Banby, "Classification of breast ultrasound images based on convolutional neural networks-a comparative study," in *Proc. 2021 International Telecommunications Conference (ITC-Egypt)*, Jul. 2021, pp. 1–8. doi: 10.1109/ITC-Egypt52936.2021.9513972

[10] M. M. Badža and M. Č. Barjaktarović, "Classification of brain tumors from MRI images using a convolutional neural network," *Applied Sciences*, vol. 10, no. 6, p. 1999, 2020. doi: 10.3390/app10061999

[11] T. M. Usman, Y. K. Saheed, D. Ignace, and A. Nsang, "Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 78–88, 2023. doi: 10.1016/j.ijcce.2023.02.002

[12] B. Kuriakose and K. P. Preena, "A review on 2D image representation methods," *Int. J. Eng. Res. Technol. (IJERT)*, vol. 4, no. 4, Apr. 2015. doi: 10.17577/IJERTV4IS041201

[13] I. E. Timor-Tritsch and A. Monteagudo, "Three-dimensional ultrasound in gynecology," *Ultrasound in Gynecology*, pp. 268–286, 2007. doi: 10.1016/B978-0-443-06630-6.50031-7

[14] E. M. Boctor, M. A. Choti, E. C. Burdette, and R. J. Webster III, "Three-dimensional ultrasound-guided robotic needle placement: an experimental evaluation," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 4, no. 2, pp. 180–191, 2008. doi: 10.1002/rcs.184

[15] S. Winzeck *et al.*, "ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI," *Front Neurol*, vol. 9, Sep. 2018. doi: 10.3389/fneur.2018.00679

[16] J. Arevalo, F. A. González, R. Ramos-Pollán, J. L. Oliveira, and M. A. G. Lopez, "Representation learning for mammography mass lesion classification with convolutional neural networks," *Comput. Methods Programs Biomed*, vol. 127, pp. 248–257, 2016. doi: 10.1016/j.cmpb.2015.12.014

[17] Y. Jia *et al.*, "Caffe," in *Proc. the 22nd ACM International Conference on Multimedia*, New York, NY, USA: ACM, Nov. 2014, pp. 675–678. doi: 10.1145/2647868.2654889

[18] Z. Jiao, X. Gao, Y. Wang, and J. Li, "A deep feature-based framework for breast masses classification," *Neurocomputing*, vol. 197, pp. 221–231, Jul. 2016. doi: 10.1016/j.neucom.2016.02.060
[19] A. Fenster and D. B. Downey, "Three-dimensional ultrasound imaging," *Physics in Medicine & Biology*, pp. 1–10, Sep. 2001. doi: 10.1117/12.440246
[20] D. Pan *et al.*, "Adaptive 3DCNN-based interpretable ensemble model for early diagnosis of Alzheimer's disease," *IEEE Trans. Comput. Soc. Syst.*, vol. 11, no. 1, pp. 247–266, Feb. 2024. doi: 10.1109/TCSS.2022.3223999
[21] G. Meena, K. K. Mohbey, and S. Kumar, "Monkeypox recognition and prediction from visuals using deep transfer learning-based neural networks," *Multimed Tools Appl.*, vol. 83, no. 28, pp. 71695–71719, 2024. doi: 10.1007/s11042-024-18437-z
[22] M. Zhu, S. Bin, and G. Sun, "Lite-3DCNN combined with attention mechanism for complex human movement recognition," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, 2022. doi: 10.1155/2022/4816549
[23] T. H. Le, T. M. Le, and T. A. Nguyen, "Action identification with fusion of BERT and 3DCNN for smart home systems," *Internet of Things*, vol. 22, 100811, Jul. 2023. doi: 10.1016/j.iot.2023.100811
[24] Y. Guo, L. Yi, and X. Pei, "3D CNN classification model for accurate diagnosis of coronavirus disease 2019 using computed tomography images," *Journal of Medical Imaging*, vol. 8, no. S1, 2021. doi: 10.1117/1.JMI.8.S1.017502
[25] S. P. Singh, L. Wang, S. Gupta, B. Gulyas, and P. Padmanabhan, "Shallow 3D CNN for detecting acute brain hemorrhage from medical imaging sensors," *IEEE Sens. J.*, vol. 21, no. 13, pp. 14290–14299, 2021. doi: 10.1109/JSEN.2020.3023471
[26] L. Yuan, X. Wei, H. Shen, L.-L. Zeng, and D. Hu, "Multi-center brain imaging classification using a novel 3D CNN approach," *IEEE Access*, vol. 6, pp. 49925–49934, 2018. doi: 10.1109/ACCESS.2018.2868813.\
[27] I. Sahumbaiev, A. Popov, J. Ramirez, J. M. Gorriz, and A. Ortiz, "3D-CNN HadNet classification of MRI for Alzheimer's disease diagnosis," in *Proc. 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, 2018, pp. 1–4. doi: 10.1109/NSSMIC.2018.8824317
[28] W. Alakwaa, M. Nassef, and A. Badr, "Lung cancer detection and classification with 3D Convolutional Neural Network (3D-CNN)," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 8, 2017. doi: 10.14569/IJACSA.2017.080853
[29] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med Image Anal.*, vol. 36, pp. 61–78, 2017. doi: 10.1016/j.media.2016.10.004
[30] C. Chen, X. Liu, M. Ding, J. Zheng, and J. Li, "3D dilated multi-fiber network for real-time brain tumor segmentation in MRI," in *Proc. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference*, Shenzhen, China, 2019, pp. 184–192. doi: 10.1007/978-3-030-32248-9_21
[31] Y. Chen, Y. Kalantidis, J. Li, S. Yan, and J. Feng, "Multi-fiber networks for video recognition," *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 364–380. doi: 10.1007/978-3-030-01246-5_22
[32] H. Lee, T. Matin, F. Gleeson, and V. Grau, "Efficient 3D fully convolutional networks for pulmonary lobe segmentation in CT images," arXiv Print, arXiv: 1909.07474, Sep. 2019.
[33] C. Yang, A. Rangarajan, and S. Ranka, "Visual explanations from deep 3D convolutional neural networks for Alzheimer's disease Classification," *AMIA Annual Symposium Proceedings*, p. 1571, 2018.
[34] K. R. Kruthika, Rajeswari, and H. D. Maheshappa, "CBIR system using capsule networks and 3D CNN for Alzheimer's disease diagnosis," *Inform Med Unlocked*, vol. 14, pp. 59–68, 2019. doi: 10.1016/j.imu.2018.12.001
[35] C. Feng *et al.*, "Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM," *IEEE Access*, vol. 7, pp. 63605–63618, 2019. doi: 10.1109/ACCESS.2019.2913847
[36] V. Wegmayr, S. Aitharaju, and J. Buhmann, "Classification of brain MRI with big data and deep 3D convolutional neural networks," in *Proc. Medical Imaging 2018: Computer-Aided Diagnosis*, 2018, pp. 406–412. doi: 10.1117/12.2293719
[37] D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," in *Proc. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference*, Athens, Greece, 2016, pp. 212–220. doi: 10.1007/978-3-319-46723-8_25
[38] A. Amidi, S. Amidi, D. Vlachakis, V. Megalooikonomou, N. Paragios, and E. I. Zacharaki, "EnzyNet: Enzyme classification using 3D convolutional neural networks on spatial representation," *PeerJ*, vol. 6, p. e4750. 2018. doi: 10.7717/peerj.4750
[39] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, "3D Deep Learning on Medical Images: A Review," *Sensors*, vol. 20, no. 18, p. 5097, 2020. doi: 10.3390/s20185097