A Contemporary CNN Based Classifier Approach for Intelligent Diagnostic Systems

[1]Swetha Parvatha Reddy Chandrasekhara, [2]Srivinay, [3]Sreevidya B S and [4]Rudramurthy V C
[1,2,3,4]Department of Information Science and Engineering, B.M.S. College of Engineering, Bengaluru, Karnataka, India.
[1]swethapc.ise@bmsce.ac.in, [2]srivinay.ise@bmsce.ac.in, [3]sreevidyabs.ise@bmsce.ac.in, [4]rudramurthyvc.ise@bmsce.ac.in

Correspondence should be addressed to Swetha Parvatha Reddy Chandrasekhara : swethapc.ise@bmsce.ac.in

Article Info
Journal of Machine and Computing (https://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202505013
Received 18 March 2024; Revised from 02 October 2024; Accepted 22 October 2024.
Available online 05 January 2025.

Abstract – Intelligent diagnostic systems significantly enhance the effectiveness and efficiency of cancer detection and management, ultimately leading to better patient outcomes. According to statistics, cancer is the second prime cause of death in males. It's a sluggish-growing ailment that doesn't show symptoms until it's quite evolved. Various investigations on AI (Artificial Intelligence) algorithms analysis have been done in the previous few years over varied medical imaging modalities which includes Computed Tomography, Magnetic Resonance Imaging, and Ultrasound. The use of artificial intelligence to monitor prostate cancer would have a tremendous impact on healthcare. Cancer scientists would have a superior understanding of the ailment and it would be helpful in developing a more precise mechanism for cancer detection as it is the need of the hour, as it has been predicted that there will be over 1.3 million additional cases diagnosed annually around the world. Here an attempt has been made to provide an analysis of the progress being made in the sector of medical image processing. Also, based on the rising interest in CNN (Convolutional Neural Networks) in recent years, we have examined the use of CNN in numerous automatic processing tasks for prostate cancer identification and diagnosis. In this study, a novel deep learning convolutional neural network (CNN) model was employed and its performance was compared against three established CNN models: AlexNet, GoogleNet, and ResNet. It has been found that the use of CNN has increased dramatically, with excellent outputs gained using either new models or pre-conditioned networks for transfer learning. Deep learning-based research surpasses traditional patient prognostic methods with regard to accuracy, according to the survey's findings.

Keywords – Cancer, Artificial Intelligence, Convolutional Neural Network, Deep Learning, Magnetic Resonance Imaging.

I. INTRODUCTION

Prostate cancer is a disease that affects the male reproductive system's prostate gland. In the US, around 209 thousand men are diagnosed with cancer each year. This illness kills 29,970 men in there. According to the society of American Cancer, one seventh of the total male population may develop prostate cancer over his lifetime [1]. This malignancy mostly affects men who are older than age of 60 Years. Early indications of this malignancy are absent, making it difficult to diagnose in its early stages. Men with advanced prostate cancer, on the other hand, have urinary issues. A digital rectal examination (DRE) was performed to see if the size and shape of the prostate gland were abnormal, and if so, a PSA test was suggested. Prostate-specific antigen is released by the prostate gland (PSA). PSA can be detected in the blood. PSA values beyond a certain threshold suggest the presence of a prostate problem. However, it is not totally definite that the high level of PSA is related to hormonal fluctuations caused by medicine that changes the hormones and causes them to create an elevated level of PSA [2]. The next step is to have a magnetic resonance imaging (MRI) or computed tomography (CT) scan. If the outcomes are positive, biopsies will be performed to establish if the tumor is benign or malignant. However, new advances in the field of neural networks have paved the path for non-invasive prostate cancer diagnosis.

Convolutional Neural Networks (CNNs), also referred to as ConvNets, are complex feedforward neural networks utilized within the field of machine learning. Renowned for their exceptional accuracy, CNNs find application in image classification and recognition tasks. Yann LeCun [3], a notable computer scientist, introduced this method of image classification in the late 1990s, drawing inspiration from human visual perception for object recognition. The architecture of CNN involves a hierarchical structure that forms a network resembling a funnel. This structure eventually leads to a fully-connected layer where all neurons are interlinked, thereby processing the resulting information.

CNNs have established themselves as the preferred choice for object recognition due to their robustness, ease of training, and manageable control. Even when employed with extensive datasets, they do not demonstrate significant levels of

overfitting. The performance of CNNs closely rivals that of equivalently sized traditional feedforward neural networks. The primary challenge, however, lies in their application to high-resolution images. Handling images of such quality necessitated a significant upheaval at the scale of ImageNet. This involved optimizing GPUs, reducing training durations, and simultaneously enhancing performance.

In the domains of computer vision and image processing, convolutional neural networks stand out as one of the most transformative advancements. The term "multilayer perceptron" often alludes to networks that possess full connectivity, signifying that each neuron in a given layer is interconnected with all other neurons in the subsequent layer. However, this fully interconnected nature renders such networks susceptible to the issue of data overfitting. To counteract this, various techniques can be employed to regularize the data, including methods like weight decay or strategies that decrease connectivity, such as skipped connections and dropout [4].

In comparison, CNNs embrace a distinct strategy for regularization: leveraging the inherent hierarchical arrangement within the data and harnessing the uncomplicated, compact patterns embedded in their filters to capture progressively intricate structures. Consequently, CNNs find their place on the lower end of the spectrum concerning connectivity and complexity. Their conception was motivated by the examination of neuron behavior within the human visual cortex.

In contradistinction to other image classification algorithms, CNNs demand minimal pre-processing. This implies that, unlike conventional methodologies, the network undergoes training to autonomously optimize its filters (or kernels) through learning mechanisms [5]. An inherent advantage stems from the fact that feature extraction doesn't rely on prior knowledge or human interaction. With deep convolutional neural networks (CNNs) maintaining their efficacy in visual recognition tasks such as object identification and segmentation, the advanced diagnostic research community is increasingly exploring various CNN models as potential contenders for developing more accurate computer-aided designs for cancer detection.

II. WORK IN THIS AREA

Mehrtash et al [6] also employed the ProstateX challenge where the dataset's ADC, increased b-value images, and Ktrans (DCE-MRI) were considered to categorize the 3D Prostate image into a cancerous lesion vs. non-cancerous lesions using 32,32, 12 ROI and a VGGNet-inspired nine convolution layers deep 3-dimension CNN classifier.

Using 172 patient’s datasets, Wang et al [7] evaluated the effectiveness of deep learning-based approaches versus non-deep learning-based approaches in classifying prostate image into cancerous MRI slices and non-cancerous MRI slices. The performance of their VGG Net-inspired 7-layer CNN classifier, which has five convolution layers and two inner product layers, was evaluated using cross-validation. They initially classified each patient's slice, then used a simple voting strategy to transform slice-level outcomes into patient-level results, yielding a patient-level AUC of 0.84, a PPV of 79 percent, and an NPV of 77 percent. It was possible to achieve similar results in our investigation by using an independent assessment set and a larger sample size.

To achieve two-dimensional (2D) ROI classification, Le et al [8] employed a mix of the merged multimodal Residual Network (ResNet)17 and the standard handmade feature extraction approach. After supplementing the training dataset, they utilized the test set to fine-tune and validate their classifier. At the ROI (lesion) level, they had an AUC of 0.91.

In 2012, Alex Krizhevesky [9] and colleagues proposed a deeper and broader CNN model than existing standard models, which achieved a futuristic recognition accuracy as compared to all conventional machine learning and computer vision approaches. AlexNet is a premier object-detection framework with a broad array of applications in the artificial intelligence area of computer vision.

Litjens et al. [10] proposed using biopsy materials with CNN to develop a model for detecting prostate cancer. The best technique to evaluate the results of CLM histogram analysis is to use ROC analysis. The average bootstrapped area under the ROC curve (AUC) for the median analysis was 99 percent, while the AUC for the 90th percentile analysis was 98 percent. The 90th percentile examination had better sensitivity than the median analysis, with a sensitivity level of 99.9%.

For the categorization of prostate cancer lesions, Liu, Zheng, Feng, and Li [11] developed a deep learning architecture called XmasNet based on Convolutional neural networks. The PROSTATEx challenge provided the system with 3D multiparametric MRI data, which it used. To integrate the 3D information about the lesion, data augmentation was done using 3D rotation and slicing.

Arvaniti et al. [12] created an Automated Gleason grading of prostate cancer tissue microarray analysis using deep learning and CNN. On the model, TMA H&E staining image patches were employed. For successfully training the classifier utilising the CNN architecture, mobile-net Transfer learning, robust regularisation, and symmetric mini-batches were proven to be crucial. When two experienced uropathologists were tested against the model, their performance accuracy was 71% and 75%, respectively.

Hu et al. [13] conducted a review of image-based diagnosis of cancer using Deep Learning. Six studies used MRI, 3D magnetic resonance volume, computed tomography slices, and histology, among other modalities, to generate CNNs. Among the applications discussed were segmentation, Lumen-based cancer diagnosis, and Gleason Grading. All of the models, with the exception of one, used end-to-end learning, while the remainder used Transfer learning.

A Deep Learning technique for enhancing Gleason grading in prostate cancer was proposed by Nagpal et al. [14]. The system consisted of two phases. The images from the dataset were fed into a CNN, which in the first phase classified each one of them as non-tumor, Gleason pattern 3, 4, or 5. The KNN classifier was employed in the second stage to determine the grade of every Gleason pattern. It was possible to achieve a 70% accuracy rate.

Goldenberg et al's [15] CNN-based model for prostatic segmentation in MRI in their review work. Prior to actually adding blocks to predict the attributes of the shape model, the network was trained to predict cancer. As a data augmentation strategy, training the photographs and reshaping its prostate surface key points according to the displacements produced from the shape model and utilizing various regularization approaches was proposed. The model had an accuracy of 88 percent and was constructed using Elastic Net architecture spectrum dropout for regularization.

Nguyen et al. [16] developed a CNN model based on U-net that correctly predicted the dose of intensity modulated radiation therapy (IMRT) for patients with prostate cancer, and also the cumulative mean and max dose differences of all structures, to within 5.1 percent of the prescribed dose. The model had a 91 % overall accuracy, and the clinical picture contours of the projected target volume (PTV) and tissues at risk (OAR) were employed as prediction parameters.

Bulten et al. [17] created an automated Deep learning system for Gleason grading using prostate biopsies. Individual glands were identified, Gleason growth patterns were assigned, and a biopsy-level grade was calculated using U-net architecture. The accuracy of the model was 91 percent. In an observation, the model outperformed a panel of pathologists, scoring 85 percent to the latter's 81%.

Strom [18] conducted a diagnostic inquiry using AI for the diagnosis and grading of prostate cancer using needle core biopsies. They used a CNN-based ImageNet architecture to evaluate the biopsies. Predicting the existence, extent, and Gleason grade of cancerous tissue was used to assess the networks. They also looked at grading and discriminating performance, as well as tumor extent estimates using anticipated cancer length correlations, using receiver operating characteristics. The area under the receiver - operating characteristic characteristics curve for the model was 99.7%. The correlation between tumor length estimated by the AI and allocated by the reporting pathologists was 96% for the autonomous test dataset and 87% for the exterior validation dataset.

III. METHODOLOGY

Basic CNN Architecture

Comprising the CNN are three distinct layer types: convolutional layers, pooling layers, and fully-connected (FC) layers [19]. Upon arranging these layers sequentially, a CNN architecture is established. Beyond these core layers, two additional crucial components come into play: the dropout layer and the activation function, both of which will be elaborated upon below.

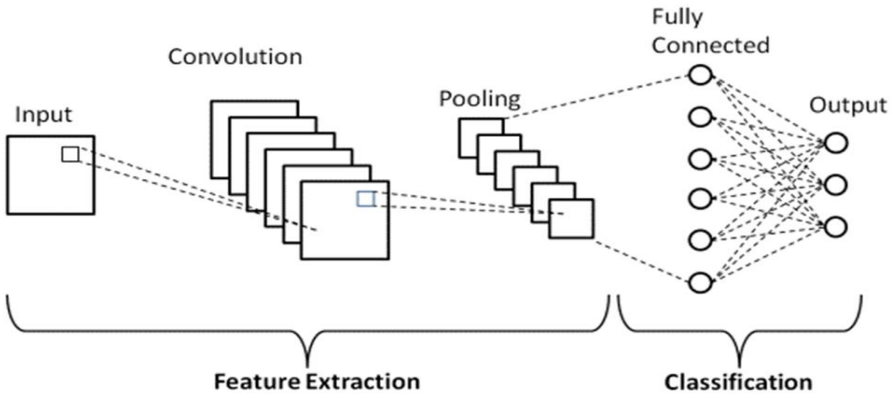


Fig 1. Convolutional Neural Network Architecture [19].

The 5 layers of the CNN are as discussed below:

Convolutional Layer

This is the initial layer, and it captures the input image's numerous features. This layer performs convolution computations between the input picture and a MxM filter. In terms of filter size, rolling the filter from across the source image produces the dot product between the filter and the sections of the source images (MxM). The Feature map is the result, and it contains information on the image's corners and edges. Other layers use the input image to learn a variety of other features, and this feature map would then be passed on to them.

Pooling Layer

A Pooling Layer is typically used after a Convolutional Layer. The primary purpose of this layer is to shrink the size of the convolved feature map with an aim to reduce the computational costs. This is accomplished by reducing the amount of links between layers and working separately on each feature map. Depending on the technique used, there are many types of Pooling procedures. In Max Pooling, the largest element is derived from the feature map. The estimate of the components in a predefined sized image segment is calculated using Average Pooling. The entire sum of the components in the predefined section is calculated using Sum Pooling. Connecting the Convolutional Layer and the Fully Connected Layer is usually done through the Pooling Layer.

Fully Connected Layer

Weights, biases, as well as neurons, make up the Fully Connected (FC) layer, that binds the neurons between two layers. The output layer of a CNN Architecture is typically positioned before the final few layers. In this stage, the preceding layers' input images are compressed and fed to the FC layer. The flattened vector is then transmitted via a few more FC layers, in which the mathematical functional operations are generally carried out. At this moment, the classification process begins.

Dropout

Once all of the parameters are connected to the Fully Connected layer, the training dataset is sensitive to overfitting issues. Overfitting occurs when a model works so well on training data that it would have an adverse influence on the performance if applied to new data. A dropout layer is used to solve this problem, which excludes a few neurons from the neural network during training, resulting in a smaller model. After a dropout of 0.3, 30% of the nodes in the neural net are dropped out at random.

Activation Functions

Finally, the activation function is one of the most crucial components of the CNN model. They're often used learn and quantify any kind of persistent and complex network variable-to-variable relationship. In simple terms, it determines which model information should be sent forward and that need not be sent at the network's end.

IV. EXISTING SYSTEM

Several cutting-edge CNN architectures have gained widespread adoption. The fundamental layers that are prevalent in most of the deep convolutional neural networks encompass the convolution layer, subsampling layer, dense layers, and the softmax layer. Typically, stacks of numerous convolutional layers and max-pooling layers are followed by fully connected and SoftMax layers in the final configuration.

Models like LeNet, VGG Net, NiN, FractalNet, and All Conv exemplify this approach. Additionally, alternative and more efficient advanced architectures have been put forth, including GoogLeNet, Residual Networks (ResNet), and AlexNet.

Although these architectures share a common foundation consisting of convolution and pooling, contemporary deep learning architectures exhibit certain topological variations. AlexNet, ResNet, and GoogLeNet have gained prominence as leading Deep Convolution Neural Network architectures due to their outstanding performance across various object recognition benchmarks.

Several of these architectures, such as GoogLeNet, AlexNet, and ResNet, are tailored for processing large-scale data, whereas the VGG network is regarded as a more general-purpose architecture [20].

AlexNet

In 2012, Alex Krizhevsky and fellow researchers introduced a CNN model that was deeper and wider compared to the prevailing standard models at that time. Their approach achieved a state-of-the-art level of recognition accuracy, surpassing all conventional machine learning and computer vision methods [21]. AlexNet, developed by Krizhevsky and team, stands out as a prominent object detection framework with versatile applications within the realm of artificial intelligence and computer vision.

The architecture of AlexNet is illustrated in **Fig 2**, encompassing a total of eight layers, specifically five convolutional layers and three fully-connected layers. However, it's not solely this attribute that sets AlexNet apart from other models.

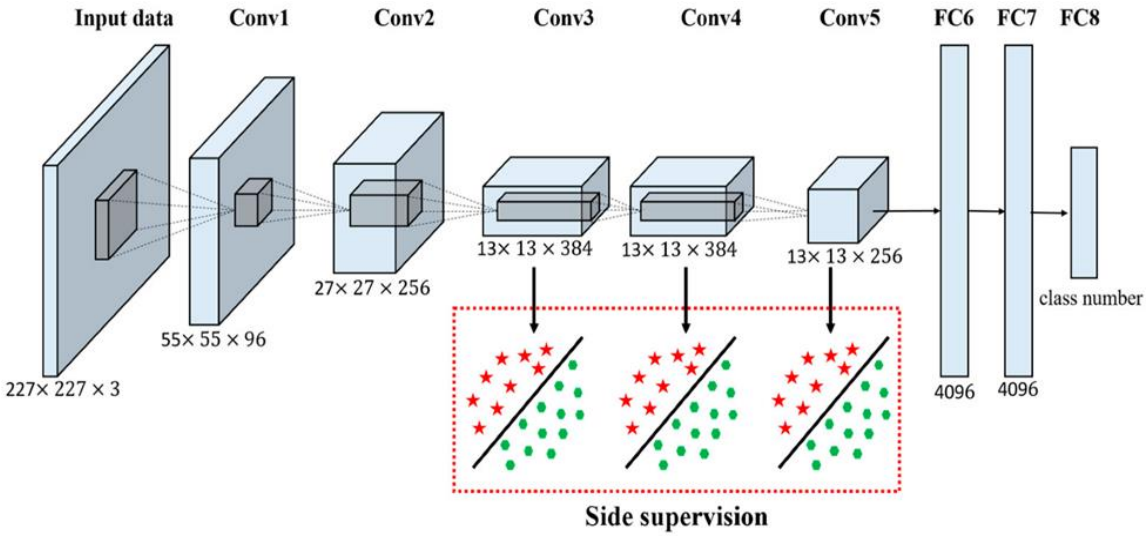


Fig 2. AlexNet Architecture [21].

There are several distinct characteristics that contribute to the uniqueness of convolutional neural networks:

ReLU Nonlinearity

AlexNet employs Rectified Linear Units (ReLU) instead of the previously prevalent tanh function. The use of ReLU confers a notable advantage in terms of training speed. Notably, a CNN incorporating ReLU achieved a 25% error rate a remarkable 60% faster compared to a CNN utilizing tanh, as demonstrated with the CIFAR-10 dataset comprising 84 images.

Multiple GPUs

In the past, GPUs were limited to a mere three gigabytes of memory. The situation was further compounded by the extensive training dataset of 1.2 million images. Addressing this challenge, AlexNet introduced the concept of multi-GPU training by partitioning half of the model's neurons onto one GPU and the remaining half onto another. This innovation not only facilitated the training of more expansive models but also significantly reduced the overall training time.

Overlapping Pooling

Within CNNs, the outputs of neighboring clusters of neurons undergo a process of "pooling" without any overlapping. Nevertheless, introducing overlap led to a notable 0.5 percent decrease in error. This observation underscores that models employing pooled layers with overlap exhibit enhanced resilience against overfitting.

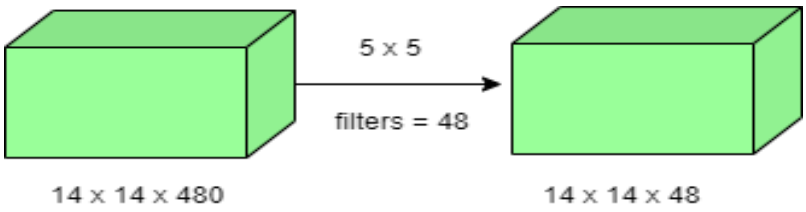
GoogLeNet

A team of Google researchers has developed GoogLeNet, a 22-layer deep convolutional neural network based on the Inception Network concept [22]. Noteworthy attributes of GoogLeNet include its marked departure from earlier state-of-the-art architectures such as AlexNet. It adopts a diverse set of techniques, including 11 convolutions and global average pooling, to achieve a deeper architectural structure. Several of these methods will be elaborated upon in the subsequent discussion of the architecture:

1×1 Convolution

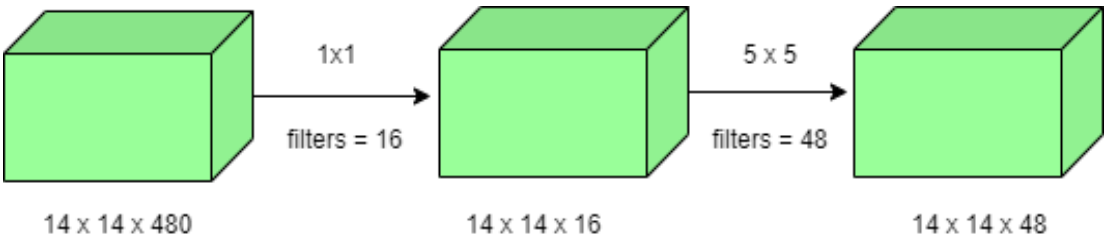
The inception architecture incorporates a total of 11 convolutions within its design. These convolutions are employed to minimize the number of weights and biases in the architecture. By reducing the parameters, the architecture's depth can be increased. To illustrate, consider the following case of a 1x1 convolution:

For instance, let's say we want to perform a 5x5 convolution with 48 filters without the need for an intermediary 1x1 convolution:



Total number of operations:(14x14x48) (5x5x480)=112.9M

With 1X1 convolution:



Total number of operations:(14x14x16) (1x1x480)+ (14x14x48) (5x5x16)=5.3M

Global Average Pooling

In prior configurations like AlexNet, fully connected layers were positioned at the network's conclusion. Across several architectures, a significant portion of parameters resided within these fully connected layers, leading to elevated computational demands. However, the GoogLeNet architecture employs a technique called global average pooling as its terminal step. This process takes a feature map with dimensions of 7x7 and averages it down to a 1x1 dimension. Consequently, the count of learnable parameters is reduced substantially, leading to an improvement of 0.6 percent in top-Accuracy.

Inception Module

The inception module contrasts with earlier architectures like AlexNet and ZF-Net due to its unique design. Within this architecture, each layer maintains a consistent convolution size. At the input stage, the Inception module concurrently performs 1x1, 3x3, 5x5 convolutions, and 3x3 max pooling. The outputs of these max-pooled layers are then amalgamated to yield the ultimate outcome. This approach theorizes that convolution filters of diverse sizes are most effective in handling objects of different scales. **Fig 3** shows Inception Module-Naïve Version [23].

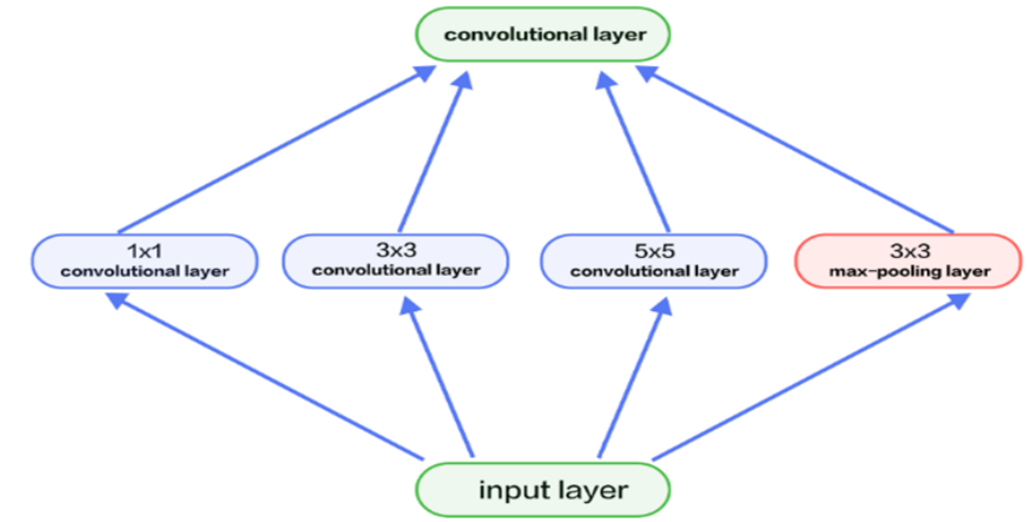


Fig 3. Inception Module-Naïve Version [23].

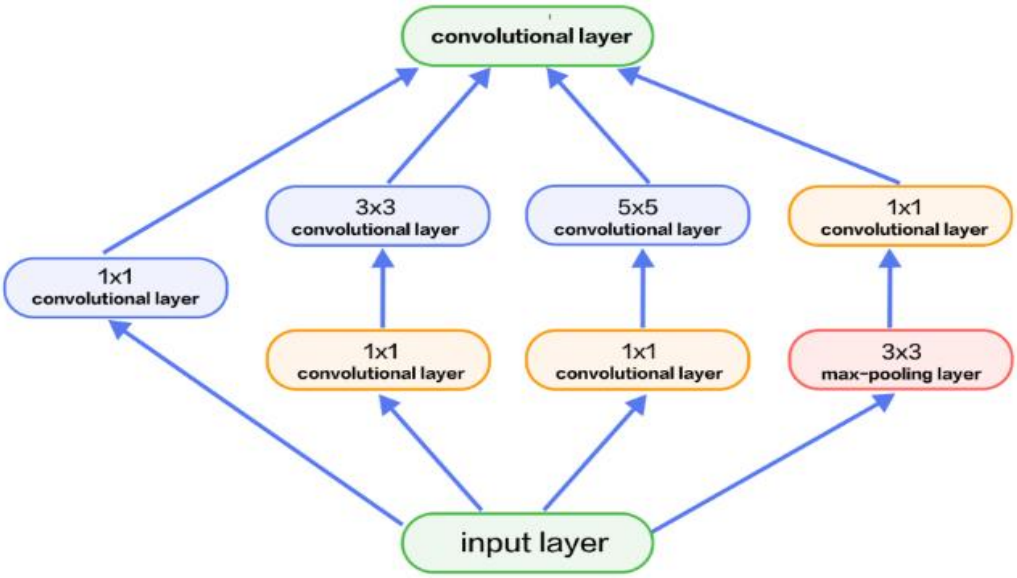


Fig 4. Inception Module- With Dimensionality Reduction [23].

Auxiliary Classifier for Training

Some intermediary classifier branches can be found in the centre of the Inception architecture. These branches are only allowed to be used for training. A 5x5 mean pooling layer with a stride of 3, 1x1 convolutions with 128 filters, fully - connected layers with 1024 and 1000 outputs, and a softmax classification layer comprise these branches. The resultant losses of these levels are then weighted and summed to the total loss. These layers aid to prevent gradient vanishing and enable regularization by preventing gradient vanishing.

Fig 5 depicts the structural framework of the GoogLeNet model. The pivotal innovation within GoogLeNet lies in its adoption of an architectural concept termed "Inception" [24]. Inception embodies a network-in-network structure and it systematically reproduces the optimal local sparse structure of a vision network from the outset to the conclusion. The architecture features three distinct Inception structures, each tailored for specific contexts. Inception relies on 1x1 convolutions as a common practice to compute reductions before engaging in more resource-intensive 3x3 and 5x5 convolutions.

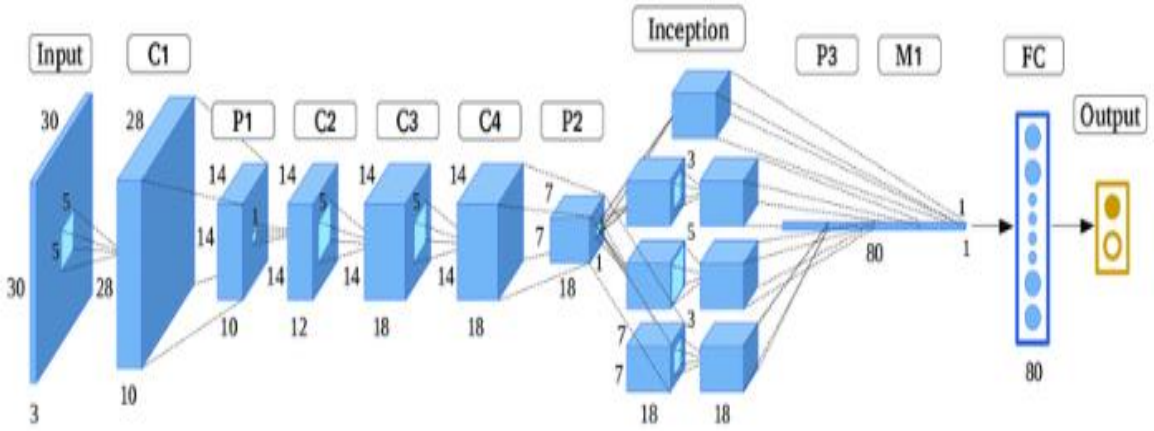


Fig 5. Architectural Outline of GoogleNet [24].

ResNet Model

A residual neural network (ResNet) is an artificial neural network (ANN) based on components observed in pyramidal cells in the cerebral cortex. Residual neural networks use skip connections, also known as shortcuts, to skip over some layers. Just between double- or triple-layer skips with nonlinearities, most ResNet models apply batch normalization (ReLU). To learn the skip weights, an extra weight matrix can be utilised; these models are known as HighwayNets.

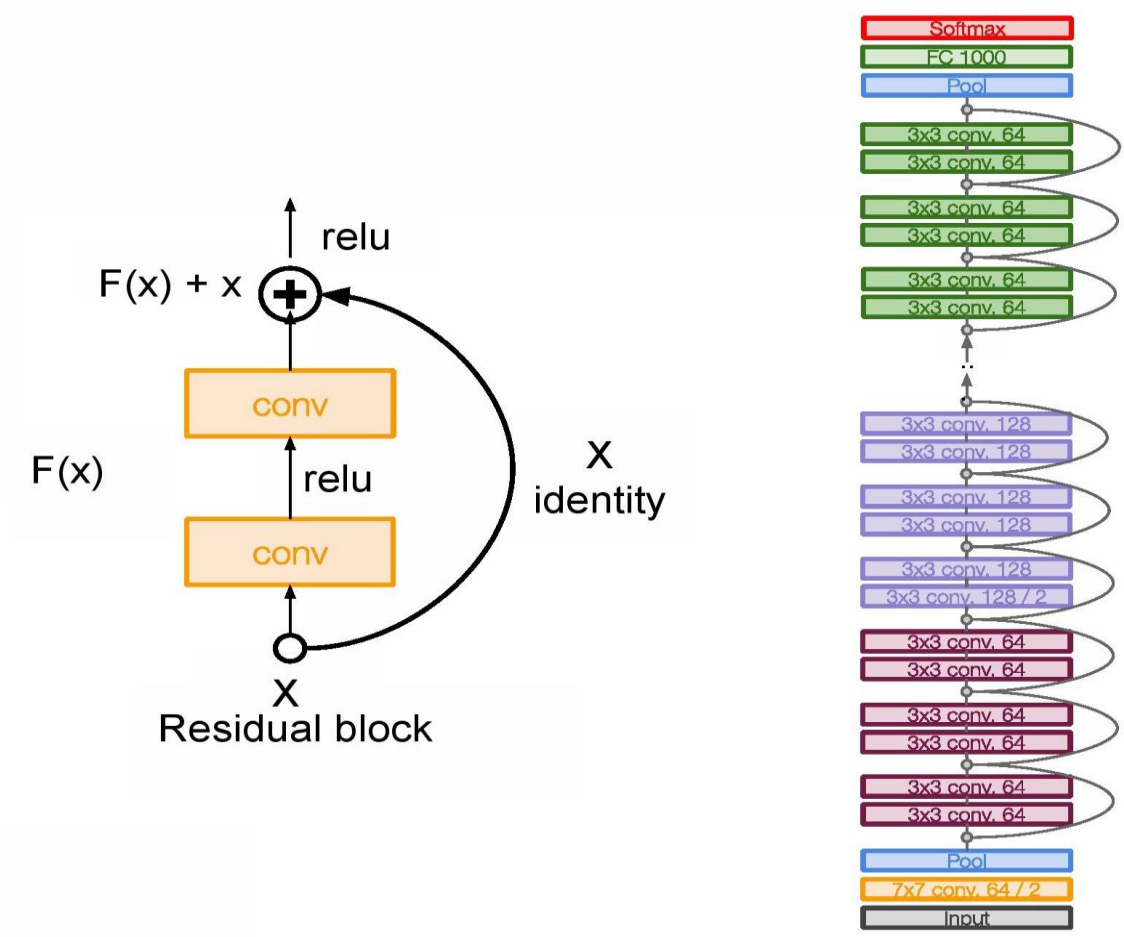


Fig 6. Resnet Architecture and Layer Details [25].

As indicated in **Fig 6**, ResNet's basic notion is to provide a so-called "identity shortcut connection" that bypasses one or more levels. The two major motivations for adding skip connections are to avoid vanishing gradients and to alleviate the Degradation (accuracy saturation) problem, which happens when introducing more layers to a sufficiently deep model leads in increased training error.

When employing gradient-based learning methods and backpropagation to train artificial neural networks, the vanishing gradient problem emerges. In such methods, each of the neural network's weight receives an update that is proportional to the partial derivative of the error function with regard to the changing weight in each iteration of training. The problem is that the gradient may be so small in some cases that the weight would be unable to modify its value.

This network uses a 34-layer simple network architecture inspired by VGG-19, wherein the bypass connection is subsequently added. As a result of these bypass connections, the design is turned into a residual network.

V. PROPOSED SYSTEM

Deep learning is a subdivision of machine learning that has evolved from previous approaches like Machine Learning to Artificial Neural Networks (ANNs). ANNs mimic the functioning of the human brain. Neurons, serving as the fundamental units of computation in ANNs, execute operations and subsequently transmit information to other neurons to perform further functions. Neurons are organized into layers; generally, computations from one layer are propagated to the following one, although certain networks enable information to flow between layers or even within neurons themselves. The ultimate outcome is generated at the output of the final layer, which can be utilized for tasks such as regression and classification. Deep learning operates by automatically extracting insights from data, facilitating the analysis and prediction of complex problems.

Deep Convolutional Neural Networks (CNNs) are a kind of Multilayer Perceptron that has excelled in a number of computer vision and image processing competitions. CNN's exciting application areas include image analysis and fragmentation, object recognition, video processing, natural language processing, and voice recognition. Deep CNN's high learning capacity is due to the use of many feature retrieval stages that can automatically learn representations from data. The availability of a large amount of data, combined with advances in hardware technology, has propelled CNN research forward, with recent reports of exciting deep CNN structures. Different activation and loss functions, parameter optimization, regularization, and architectural improvements are just a few of the novel approaches that have been investigated to help CNNs achieve breakthroughs. The deep CNN's significant increase in representational capability is due to architectural advances. Utilizing spatial and channel data, architecture depth and width, and multi-path data processing, in particular, have received a lot of attention.

This research introduces a robust deep learning convolutional neural network (CNN) known as the PyNet5 model. The approach involves transfer learning, coupled with a slight modification in the depth and breadth of the conventional CNN architecture. Subsequently, the outcomes are juxtaposed with three leading-edge CNN models, namely AlexNet, GoogleNet, and ResNet, all using the identical dataset. An overview of the architectural structure of the proposed PyNet5 model is illustrated in **Fig 7**.

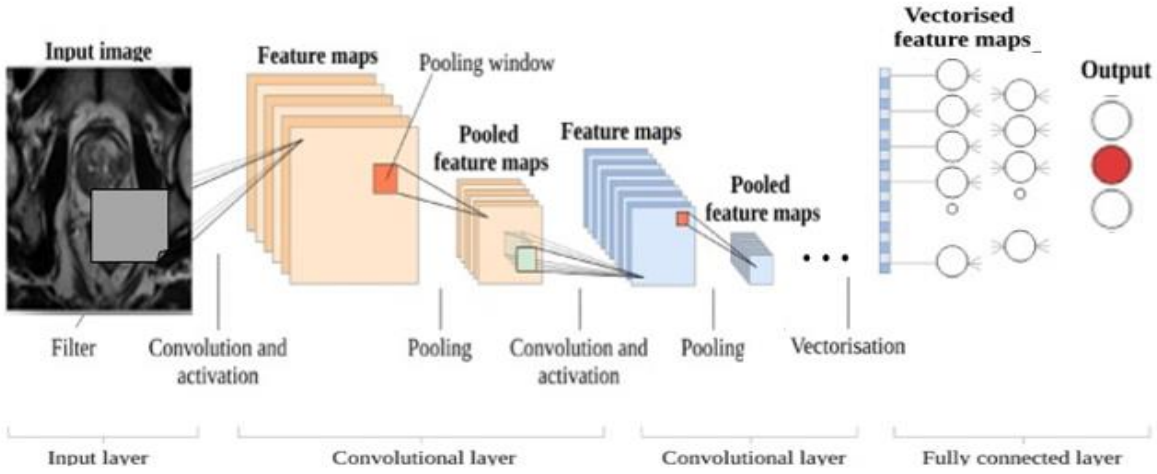


Fig 7. Architecture of proposed PyNet5 Model.

Proposed Model Framework

The model presented comprises a grand total of 8 layers, encompassing the initial input layer and the concluding output layer, while also incorporating 5 hidden layers and a fully connected layer. The subsequent sections provide a detailed breakdown of the distinct layers within the proposed PyNet model.

Input Layer

For this investigation, a grayscale MRI image is taken as the input image. Within the input layer, the image is transformed into a 200 x 200 matrix, containing a single image channel. Subsequently, a segment of the selected input image's matrix is extracted and fed into the convolutional layer, where a set of specific filters is applied to this image matrix patch.

Convolutional Layer

Within the convolutional layer, every neuron functions as a kernel, forming a collection of convolutional kernels. The image is partitioned into smaller sections known as receptive fields using these convolutional kernels. Fragmenting the image into these smaller units facilitates the extraction of feature patterns. Through multiplication with the corresponding elements of the receptive field, the kernel convolves with the images, applying a predetermined set of weights. The process of CNN filtering to generate a convolved matrix is illustrated in **Fig 8**.

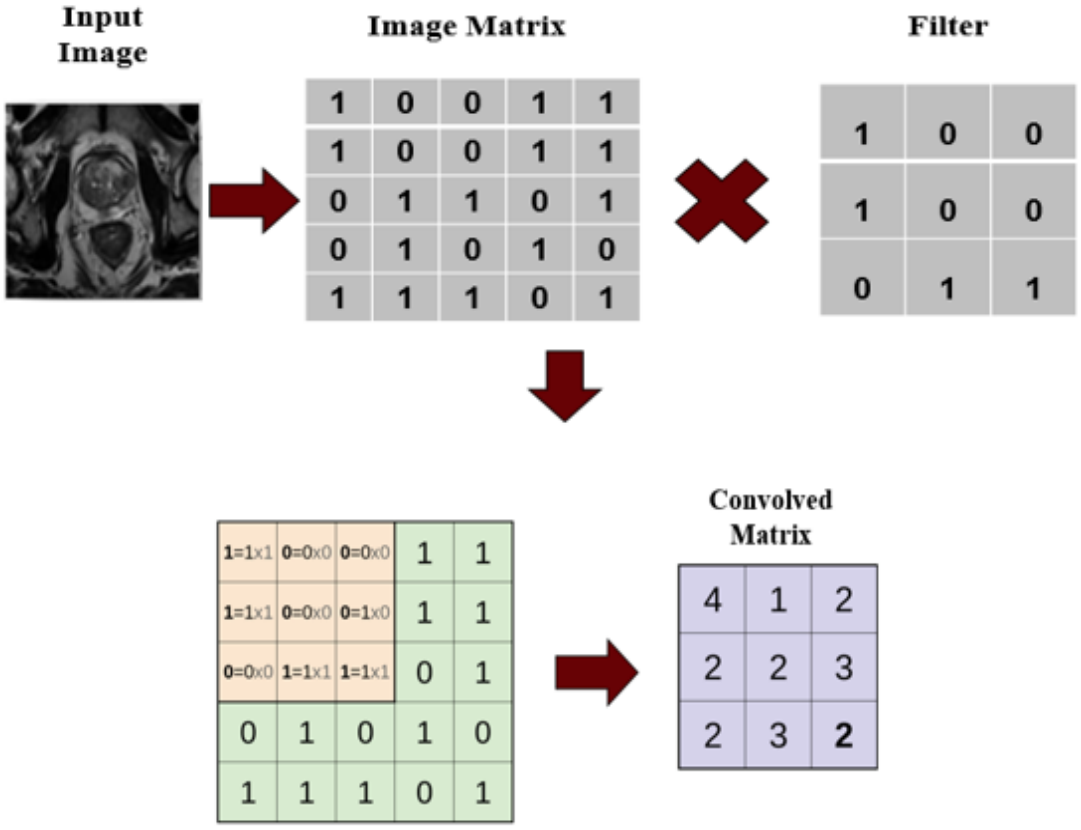


Fig 8. Convolved Matrix Using CNN Filtering.

Convolution operation is expressed as follows:

$$F_l^k(p,q) = \sum_c \sum_{x,y} i_c(x,y) \cdot e_l^k(u,v) \quad (1)$$

Where,

F_l^k : Input feature matrix for l^{th} layer and k^{th} kernel

i_c : Tensor image

$f_l^k(p,q)$: (p, q) element of feature matrix

$i_c(x,y)$: (x, y) element of c^{th} channel of an image

$e_l^k(u,v)$: (u, v) element of k^{th} kernel of l^{th} layer

Pooling Layer

The pooling layer serves as a non-linear down-sampling mechanism, reducing the image dimensions by half. Consequently, the identical image patch that underwent convolution is also subjected to the pooling layer. Feature patterns resulting from the convolution process might manifest at diverse positions within the image. After obtaining these features, their precise locations become less crucial, as long as their relative positions with respect to one another are preserved. Pooling, also referred to as down-sampling, presents an intriguing localized operation. **Table 1** shows Architectural Details of the Proposed Model.

$$z_l^k = g_p(F_l^k) \quad (2)$$

Where,

z_l^k: pooled feature-map of l^{th} layer for k^{th} input feature-map

g_p : Pooling Operation

F_l^k : Input feature matrix for l^{th} layer and k^{th} kernel

Table 1. Architectural Details of the Proposed Model

Architectural Details of PyNet5 Model		
Batch size	16	Batch size pertains to the quantity of input images that are simultaneously fed into the convolutional layer.
Epochs	5	Tells the number of iterations over the dataset in order to train the neural network.
Image Matrix	200*200	Specifies the dimensions of the matrix used for applying operations on the image.
Image channel	1	The image channel is configured as 1 in this context, as a grayscale image is chosen as the input. However, if an RGB image were to be used, the channel size would be set to 3.
Size of Filters	32	A filter size of 32 is designated for executing the convolution operation.
Hidden Layers	5	This architecture consists of a total of 5 hidden layers, comprising 3 convolutional layers and 2 pooling layers.
Number of classes	3	The output layer classifies the input image into one of three distinct categories: normal, benign, and malignant.

Fully Connected Layer

This layer is characterized by each node being directly linked to the subsequent node in a dense manner. It operates as a data-intensive node, requiring a substantial number of coefficients to support every node within the pooling layer. Typically deployed for classification towards the network's culmination, the fully connected layer represents a comprehensive operation, distinct from pooling and convolution. It accumulates data from the preceding feature extraction phases and evaluates the collective output from all prior layers. Ultimately, the fully connected layer provides the highest probability for the identified object.

Output Layer

The SVM classification technique applied in the fully connected layer yields the ultimate output comprising three distinct classes.

VI. EXPERIMENTAL ANALYSIS & RESULTS

The research was conducted using a dataset of 700 patient records, consisting exclusively of preprocessed MRI images. For the purpose of prostate segmentation, the MRI images underwent training for 5 epochs. Each of the 700 images was processed and converted into an image matrix, which was subsequently fed into the convolutional layer. During this process, the image matrices were dispatched in batches of 16 images per iteration. As the images in question were grayscale, a single channel was selected for analysis.

Once the image matrix is directed into the convolutional layer, it undergoes convolution using a filter size of 32. Subsequently, the convolved image proceeds to the pooling layer, where it is resized to alleviate the computational burden that could arise from managing extensive data volumes. Within the pooling layer, an activation function called ReLU (Rectified Linear Unit) is employed on the image's neurons. The inclusion of ReLU in the pooling layer serves the purpose of preventing simultaneous activation of all neurons. In essence, ReLU is utilized to curb the exponential surge in computations needed to operate the neural network.

Ultimately, after undergoing three convolutional layers and two pooling layers, the collection of images progresses to the fully connected layer. Within this fully connected layer, a multiclass SVM classifier is applied to categorize the images into three distinct classes: normal, benign, and malignant. A comparative performance analysis of the proposed system alongside three established CNN models is presented in **Table 2**.

Table 2. Performance Of Proposed Model Against The Existing Models

	Accuracy	Sensitivity	Specificity
AlexNet	95.73	96.63	96.99
GoogleNet	97.19	97.19	96.65
ResNet	96.37	96.37	95.93
Proposed PyNet5	99.41	99.59	99.26

Performance Assessment Was Conducted Using An Identical Dataset Across The Four CNN Models
AlexNet, GoogleNet, ResNet, and the newly introduced PyNet5. Graphs depicting the performance of these four models in accuracy, sensitivity, and specificity are illustrated in **Figs 9, 10, and 11**. The empirical findings distinctly indicate that the proposed CNN model surpasses existing state-of-the-art CNN models, attaining an accuracy, sensitivity, and specificity of 99.41%, 99.59%, and 99.26% respectively.

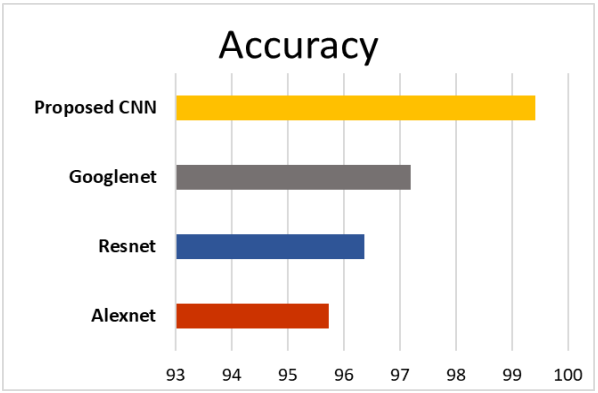


Fig 9. Accuracy of the Proposed Model in Comparison to The Existing Models.

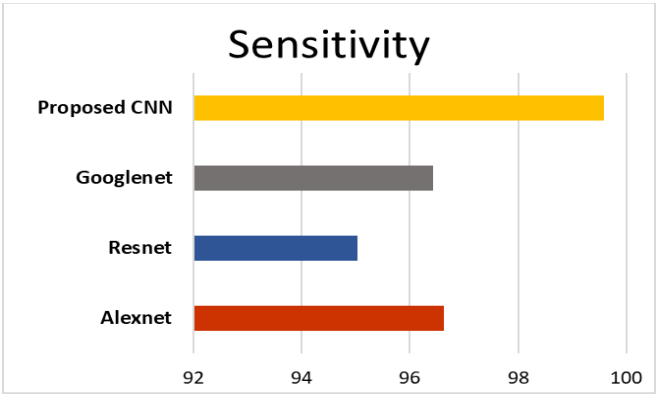


Fig 10. Sensitivity of the Proposed Model in Comparison to The Existing Models.

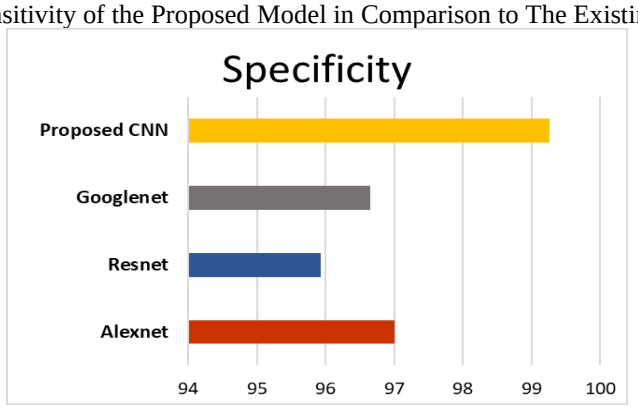


Fig 11. Specificity of the Proposed Model in Comparison to The Existing Models.

Performance Evaluation Was Conducted Using an Identical Dataset Across All Four CNN Models
AlexNet, GoogleNet, ResNet, and the novel PyNet5. Graphs represented by **Figs 9, 10, and 11** succinctly illustrate the performance of these four models in relation to accuracy, sensitivity, and specificity. The empirical findings unequivocally

demonstrate that the proposed CNN model surpasses the existing state-of-the-art CNN model, achieving accuracy, sensitivity, and specificity rates of 99.41%, 99.59%, and 99.26% respectively.

VII. CONCLUSION

In this study, a novel deep learning convolutional neural network (CNN) model was employed and its performance was compared against three established CNN models: AlexNet, GoogleNet, and ResNet. The investigation was conducted on a dataset comprising 700 patient records, all sourced from a publicly available database. All four CNN models underwent performance assessment with the identical dataset, revealing that the proposed model outperformed the other three existing models. This advantage can be attributed to inherent limitations within each state-of-the-art model. GoogleNet, for instance, suffered from a heterogeneous topology that necessitated customization from module to module. Furthermore, it exhibited a representation bottleneck that reduced the feature space in subsequent layers, occasionally leading to the loss of valuable information. On the other hand, overfitting emerged as a prominent concern for the other two models, namely AlexNet and ResNet. The superior performance achieved by the proposed model can be attributed to a substantial reduction in the number of hidden layers, which contributes to its resilience against overfitting—a notable issue in the other three models. The classification accuracy in this study was attained through the utilization of the multiclass SVM classification algorithm.

CRediT Author Statement

The authors confirm contribution to the paper as follows:
Conceptualization: Swetha Parvatha Reddy Chandrasekhara, Srivinay, Sreevidya B S and Rudramurthy V C; **Methodology:** Swetha Parvatha Reddy Chandrasekhara and Srivinay; **Software:** Srivinay, Sreevidya B S and Rudramurthy V C; **Data Curation:** Srivinay and Sreevidya B S; **Writing- Original Draft Preparation:** Swetha Parvatha Reddy Chandrasekhara and Srivinay; **Writing- Reviewing and Editing:** Swetha Parvatha Reddy Chandrasekhara and Srivinay; All authors reviewed the results and approved the final version of the manuscript.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests

References

[1]. S. P. R. Chandrasekhara, M. G. Kabadi, and S. Srivinay, “Wearable IoT based diagnosis of prostate cancer using GLCM-multiclass SVM and SIFT-multiclass SVM feature extraction strategies,” International Journal of Pervasive Computing and Communications, vol. 20, no. 1, pp. 19–37, Sep. 2021, doi: 10.1108/ijpcc-07-2021-0167.
[2]. S. P. R. Chandrasekhara, M. G. Kabadi, and Srivinay, “A Novel SIFT-SVM Approach for Prostate Cancer Detection,” Journal of Computer Science, vol. 16, no. 12, pp. 1742–1752, Dec. 2020, doi: 10.3844/jcssp.2020.1742.1752.
[3]. Y. Bengio, Y. Lecun, and G. Hinton, “Deep learning for AI,” Communications of the ACM, vol. 64, no. 7, pp. 58–65, Jun. 2021, doi: 10.1145/3448250.
[4]. J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2015, doi: 10.1109/cvpr.2015.7298965.
[5]. K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, doi: 10.1109/cvpr.2016.90.
[6]. A. Mehrtash et al., “Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks,” Medical Imaging 2017: Computer-Aided Diagnosis, vol. 10134, p. 101342A, Mar. 2017, doi: 10.1117/12.2277123.
[7]. S. Wang, K. Burtt, B. Turkbey, P. Choyke, and R. M. Summers, “Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research,” BioMed Research International, vol. 2014, pp. 1–11, 2014, doi: 10.1155/2014/789561.
[8]. M. H. Le et al., “Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks,” Physics in Medicine & Biology, vol. 62, no. 16, pp. 6497–6514, Jul. 2017, doi: 10.1088/1361-6560/aa7731.
[9]. Sergey Ioffe and Christian Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” Proceedings of the 32 nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP Vol 37, pp. 448– 456.
[10]. G. J. S. Litjens, P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman, “Automatic computer aided detection of abnormalities in multi-parametric prostate MRI,” Medical Imaging 2011: Computer-Aided Diagnosis, vol. 7963, p. 79630T, Mar. 2011, doi: 10.1117/12.877844.
[11]. S. Liu, H. Zheng, Y. Feng, and W. Li, “Prostate cancer diagnosis using deep learning with 3D multiparametric MRI,” Medical Imaging 2017: Computer-Aided Diagnosis, vol. 10134, p. 1013428, Mar. 2017, doi: 10.1117/12.2277121.
[12]. E. Arvaniti et al., “Automated Gleason grading of prostate cancer tissue microarrays via deep learning,” Scientific Reports, vol. 8, no. 1, Aug. 2018, doi: 10.1038/s41598-018-30535-1.
[13]. Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, and Q. Sun, “Deep learning for image-based cancer detection and diagnosis − A survey,” Pattern Recognition, vol. 83, pp. 134–149, Nov. 2018, doi: 10.1016/j.patcog.2018.05.014.

[14]. K. Nagpal et al., “Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer,” npj Digital Medicine, vol. 2, no. 1, Jun. 2019, doi: 10.1038/s41746-019-0112-2.
[15]. S. L. Goldenberg, G. Nir, and S. E. Salcudean, “A new era: artificial intelligence and machine learning in prostate cancer,” Nature Reviews Urology, vol. 16, no. 7, pp. 391–403, May 2019, doi: 10.1038/s41585-019-0193-3.
[16]. D. Nguyen et al., “A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning,” Scientific Reports, vol. 9, no. 1, Jan. 2019, doi: 10.1038/s41598-018-37741-x.
[17]. W. Bulten et al., “Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study,” The Lancet Oncology, vol. 21, no. 2, pp. 233–241, Feb. 2020, doi: 10.1016/s1470-2045(19)30739-9.
[18]. P. Ström et al., “Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study,” The Lancet Oncology, vol. 21, no. 2, pp. 222–232, Feb. 2020, doi: 10.1016/s1470-2045(19)30738-7.
[19]. S. G. Armato, N. A. Petrick, and K. Drukker, “PROSTATEx: Prostate MR Classification Challenge (Conference Presentation),” Medical Imaging 2017: Computer-Aided Diagnosis, p. 274, Apr. 2017, doi: 10.1117/12.2280374.
[20]. J. Glaister, A. Cameron, A. Wong, and M. A. Haider, “Quantitative investigative analysis of tumour separability in the prostate gland using ultra-high b-value computed diffusion imaging,” 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 420–423, Aug. 2012, doi: 10.1109/embc.2012.6345957.
[21]. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” OSDI'16: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, 2016, pp. 265-283.
[22]. A. Gertych et al., “Machine learning approaches to analyze histological images of tissues from radical prostatectomies,” Computerized Medical Imaging and Graphics, vol. 46, pp. 197–208, Dec. 2015, doi: 10.1016/j.compmedimag.2015.08.002.
[23]. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, doi: 10.1109/cvpr.2017.243.
[24]. W. Li et al., “Path R-CNN for Prostate Cancer Diagnosis and Gleason Grading of Histological Images,” IEEE Transactions on Medical Imaging, vol. 38, no. 4, pp. 945–954, Apr. 2019, doi: 10.1109/tmi.2018.2875868.