ROBUST FEATURE BASED CANCER CLASSIFICATION

Abstract

Brain cancer has a very short life expectancy. Accurately diagnosing brain cancer is an important step in developing an appropriate treatment plan for brain cancer treatment and rehabilitation. Computer-based cancer detection systems and convolutional neural networks have created success stories and machine learning-based disease detection has made significant progress. The present work fills this gap. Here, we used a statistics-based feature-extraction method and machine learning-based extraction method to distinguish cancer genes from non-cancer genes. Deep learning algorithms have also been used to develop automated cancer cell prediction systems. Unlike available methods, three function output modules were developed. The focus of this paper is to identify genetic hallmarks of cancer. In this proposed work, convolutional neural networks are established to predict cancer cells. Histopathological images are initially splitted into train, validation and test data in the ratio 75:15:10. Next, we ran all training on TensorFlow using Python's Keras library. The developed method proved its effectiveness on the derived function by achieving 99% accuracy. The CNN model was compared with existing work and achieved 99% accuracy. The process of binary classification as bad or good is relatively short. Here, we present an automatic binary classification method using deep learning for brain cancer classification. A deep neural network model is developed and applied as a feature extractor to learn the histopathological images and their convolutional layer structures on the histopathological image dataset. Then, the fully connected layers are replaced and the entire deep network is trained as a classifier to distinguish the two cancer classes. Our

Authors

Indranil Banerjee
NSHM Knowledge Campus
Durgapur, West Bengal, India.
banerjeeindranil82@gmail.com

Dr. Bijoy Kumar Mandal
NSHM Knowledge Campus
Durgapur, West Bengal, India.
writetobijoy@gmail.com

proposed model has about 4.2 million training parameters, which is less than VGG19, ResNet50 and InceptionV3, which have 20.1 million, 23.9 million and 21.8 million, respectively**.**

Keywords: A deep neural network model is developed and applied as a feature extractor to learn the histopathological images and their convolutional layer structures on the histopathological image dataset.

I. INTRODUCTION

At present, there is a growing trend in both the occurrence and fatality of cancer, positioning it as a prominent contributor to global mortality. According to data from 2018 [1], there were approximately 181 million new cases of cancer and 96 million cancer-related deaths. Among various types of cancer, lung cancer stands out with the highest mortality rate, closely trailed by colon cancer. Recent findings from cancer research articles [2] highlight colorectal and liver cancers. Notably, breast cancer emerges as the primary factor behind fatalities linked to lung, liver, colon, and stomach cancers on a global scale. One eighth women will develop breast cancer, which is shocking. Cancer risk factors (environment, diet, age, etc.) are not easily controllable, and cancer prevention and research are still gaps. World Health Organization surveys and studies [3] conclude that cancer patients should be diagnosed in the preliminary stages of the disease. Early detection is important as chance of successful treatment is close to 80%.

The stages of injury diagnosis include physical examination, diagnostic medical imaging (DMI) [4], clinical examination, biological examination, etc. Imaging (ultrasound , radio, various contrast agents, nuclides, CT Scan, magnetic resonance imaging (MRI) [5]) is the best method used by most scientists, but histopathological imaging[6,7] is the clinical image that is obtained from biopsy tissue. Tissues can be studied under a microscope to obtain information about cell movement. Histopathological methods are considered the best standard for the classification and diagnosis of cancer. When observing tissue through a microscope, all the structures seem uniformly shaded in dark brown, which makes it challenging to discern structural details unless they are highlighted with color. Hematoxylin and Eosin (H&E) staining is a widely employed method in histology labs for this purpose. Through H&E staining, nuclear chromatin and nucleic acids within the cytoplasm are colored, while cytoplasmic and extracellular matrix elements take on a purple hue. Additionally, nuclear chromatin within the cytoplasm appears red, contributing to enhanced data precision.

Histopathological imaging has the benefit of freely magnification which clearly shows the microscopic details in the shape of the tissue. Enhancements will be made to increase the precision of detection. In routine cancer screening, the pathologist directly observes, analyzes, and evaluates tissue under the microscope. This is a broadly accepted criterion for scientific diagnosis of cancer however, pathologists often experience faintness and lathergy under the microscope. Simultaneously, the trainee is required to document and note down the ultimate diagnosis, a process that demands a significant amount of time. Given the substantial rise in the prevalence of life-threatening illnesses, the manual examination of joints for Ayurvedic practitioners becomes exceedingly time-consuming and labor-intensive. Furthermore, manual detection can result in complications, internal irregularities, and unidentified tracking inconsistencies. There is a 20% difference in cancer diagnosis [8] between experienced and novice pathologists. A recent study found a 75.3% discrepancy between individual diagnoses and technical diagnoses [9]. Therefore, a detection method based on artificial vision is needed. It can perform efficient, stable and accurate digital image analysis. Machine vision technology overcomes the conflict between internal and external observers to increase the accuracy and consistency of cancer diagnosis and treatment.

Machine vision technology offers the benefits of remote sensing and image analysis, addressing the limitations of conventional scanning techniques and enhancing decision-

making processes. In contrast to manual detection, machine vision technology stands out for its efficiency and precise focus. The emergence of deep learning [10] in recent years has triggered a comparison between traditional machine vision approaches and deep learning methodologies. While traditional artificial vision techniques [11] offer numerous effective methods, they require manual data labeling, a complex process that often leads to unstable detection accuracy. The complexity of classification and identification increases as the number of categories expands. Deep learning exhibits substantial cognitive advantages over traditional machine vision technologies. It can achieve superior image quality and attain heightened accuracy in tasks such as image classification [12], semantic segmentation [13], and target detection [14]. However, deep learning heavily relies on extensive labeled data for goal definition and might not be universally adaptable [15]. Combining deep learning with conventional methods can potentially surmount learning challenges, leveraging in-depth knowledge, computational resources, time, textual information, and energy processing.

This research delves into machine vision methodologies designed to detect tumor cells within histopathological samples. It encompasses tasks such as image preprocessing, segmentation, and the extraction and presentation of image features using machine vision techniques. Numerous evaluations exist regarding the application of machine vision for the identification of malignancies in histopathological analyses.

II. LITERATURE REVIEW

Classification of brain tumors [15] requires identification of variables such as shape and texture in histopathological images. The most common method for detecting this cancer, currently popular among computer vision researchers, is computed tomography (CT). In a CT object, there are two main parts involved. The first is to treat and detect cancer at an first stage and the second is to classify the cancer appropriately. The research focuses on the role of brain tumors considering the author's previous works on a particular topic.

According to a study conducted by Yashwant Kurmi, Vijayshri Chaurasia and Narayanan Ganesh in the year 2019 [16], the proposed method classifies images based on hand drawn elements and textures using Bag of Visual Words (BoW). The method of segmentation of multiple nuclei in histopathology images involves decomposition of spots and scaling of histograms to confirm nuclear area, followed by radial symmetry, combined slices and controlled by a function of rapid changes in nuclear boundary based "nucleus key points" is included. Locally obtained parts were extracted to modify the manufactured parts and components used by BoW were used to capture the utility of local features and global features. A comparison was carried out on the Bisque and BreakHis datasets (with an average accuracy of 93.87% and 96.96%) and the improved analysis performance of the developed method was confirmed.

According to a study conducted by Khan MA, Ashraf I, Alhaisoni M and many more in the year 2020 [17] the clustering method used few BRATS datasets, BRATS2012, BRATS2013, BRATS2015, BRATS2018 and achieved good results. A binary classification approach proves inadequate. In the initial phase, linear smoothing techniques were implemented, employing histogram edge classification and the discrete cosine transform (DCT). Subsequently, deep learning methodologies were harnessed in the second stage. Utilizing two neural network models, VGG16 and VGG19, transfer learning was employed for extracting features. Moving forward, a hierarchy-based classification technique involving

the Edge Learning Machine (ELM) was adopted to select the most effective attributes. Finally, a consolidation of partial least square (PLS) models yielded a unified matrix. The integration matrix represents the ultimate classification outcome of ELM. This proposed method's efficacy was substantiated through validation on the BraTS dataset, resulting in accuracy rates of 97.8% for BraTs2015, 96.9% for BraTs2017, and 92.5% for BraTs2018.

Another study was conducted by Adeola Adebanjo, Qasem Abu Al-Haija in the year 2020 [18], proposed a system for breast cancer diagnosis with ResNet-50 neural network to classify histopathological images. The proposed model uses the ResNet-50 CNN to transfer robust techniques previously trained on ImageNet to classify the BreakHis datasets as good or bad. Simulation results show that the model outperforms comparable models trained on the same data set, achieving up to 99% classification accuracy.

Another study by Marriam Nawaz, Zahid Mehmood, and some more authors in the year 2021 [19], an initial introduction outlines a fully automated technique for the segmentation of skin melanoma, utilizing a deep learning-based strategy known as region-based accelerated convolutional neural networks (RCNNs) combined with fuzzy mean clustering (FKM). This approach aims to enhance the identification of this life-threatening condition by employing diverse clinical images, aiding dermatologists in early detection. The method begins by preprocessing dataset images to eliminate noise and address illumination issues, subsequently refining visual data through the acquisition of fixed-length feature vectors prior to the application of a more potent RCNN. Subsequently, FKM is employed to partition melanoma-affected skin into distinct regions based on size and boundaries. The method's efficacy is assessed across three benchmark datasets - ISBI-2016, ISIC-2017, and PH2. The outcomes highlight the superiority of the proposed approach over existing techniques. Across the mentioned datasets, the proposed method attains an average accuracy of 95.40% for ISIC-2016, 93.1% for ISIC-2017, and 95.6% for the PH2 dataset.

According to Naresh Kumar, Manoj Sharma and some more authors in the year 2022 [20], this paper presents a comparative analysis of two distinct extraction methodologies. The first approach introduces six manual drawing techniques based on color, shape, and texture attributes. These techniques were incorporated into the process. Subsequently, classifiers including SVM-RBF, Gradient Boosting (GB), Multilayer Perceptron (MLP), and Random Forest (RF) were trained and examined for the classification of colon and lung cancers. In contrast, the second approach employed the concept of transfer learning. This involved selecting seven deep features to identify abnormalities in histopathological images of gastric and lung cancer. These deep features were employed as inputs for standard classifiers including SVM-RBF, GB, RF, and MLP. Notably, a substantial performance enhancement was observed with the utilization of deep CNN-derived features compared to the manual ones. The findings demonstrated the effectiveness of the proposed method in terms of specificity, recall, F1 score, and ROC-AUC. Particularly, the RF classifier using DenseNet-121 derived features achieved an accuracy of 98.60% for lung cancer and 98.63% for colon cancer classification. Furthermore, it attained a recovery rate of 98.63%, an F1 score of 0.985, and a notable ROC-AUC value.

A recent study conducted by V. P. Subramanyam Rallabandi, Mahati Munikoti Srikantamurthy and some more authors in the year 2023 [21], a hybrid model was meticulously crafted for the purpose of categorizing four distinct breast cancer types and four variants of invasive tumors. This model amalgamates the capabilities of Convolutional

Neural Networks (CNNs) and Long Short-Term Memory Recurrent Neural Networks (LSTM RNNs). The proposed hybrid, termed CNN-LSTM, utilizes ImageNet and employs a permutation technique to capture and evaluate the four facets within each class. To gauge its effectiveness, the proposed model was assessed using the BreakHis dataset, which comprises 2,480 images of solid tumors and 5,429 images of malignant tumors across magnifications of 40x, 100x, 200x, and 400x. The hybrid model, denoted as CNN-LSTM VGG-16, was pitted against existing CNN models such as ResNet50 and the original models employed for classification. All the models were trained using common optimizers like Adam, with the inclusion of three algorithms: Root Mean Square Propagation (RMSProp) and Stochastic Gradient Descent (SGD). Through the findings, it becomes evident that the Adam optimizer stands as the most optimal choice, yielding superior loss models during both training and validation stages. The hybrid CNN-LSTM model, as presented in this study, achieved an impressive accuracy of 99% for the two-dimensional representation of yellow and green tumors. These categories collectively accounted for 92.5% of all cases of yellow-green cancers.

According to Mangal Sanidhya , Chaurasia Aanchal and Khajanchi Ayush in the year 2020 [22], histological diagnosis plays a pivotal role in determining the specific type of cancer. This study aims to introduce a neural network-based approach for the diagnosis of squamous cell carcinoma, lung adenocarcinoma, and colon adenocarcinoma. Consequently, artificial intelligence (AI) emerges as a fundamental technology in shaping the future of cancer diagnosis. The LC25000 dataset was leveraged, consisting of 2,500 digital images, with each class containing 5,000 images. Employing a deep neural network model, the study focused on classifying histopathological images related to squamous cell carcinoma, adenocarcinoma, and lung cancer. Additionally, the model was employed to differentiate adenocarcinoma from benign conditions in the colon. Remarkably, the diagnostic accuracy achieved noteworthy results, reaching 97% for lung cancer and 96% for colon cancer. These outcomes underline the potential of neural networks in enhancing the precision of cancer diagnosis within histopathological contexts.

A study conducted by Maha Mohammed Khan, Sumaiya Dabeer, Saiful Islam in the year 2019 [23], breast cancer, a condition affecting 1 in 8 women globally, is characterized by the deterioration of breast tissue. Contemporary medical imaging employs histological images obtained through microscopy, subjected to analysis using an array of algorithms and techniques. Machine learning has now entered the realm of medical imaging and diagnostics, revolutionizing their development. Manual diagnosis of melanoma remains a challenging undertaking, often prone to human errors. Consequently, the employment of manual methods can yield superior outcomes in contrast to conventional testing procedures. In the domain of deep learning, this is accomplished through the utilization of artificial neural networks (ANNs) to extract features and subsequently classify them via a fully connected network. The integration of deep learning has found widespread application in the field of medical imaging, primarily due to the simplicity of executing analogous tasks, sans the requirement of intricate technology. In this study, an adaptive neural network was trained, culminating in an impressive predictive success rate of 99.86%. This underscores the potential of such technologies in advancing accurate medical predictions and diagnoses.

Another study by Babu, Tina in the year 2021 [24], an automated pancreatic cancer diagnostic system has been developed, employing cross-learning to automatically extract high-resolution features from pancreatic biopsy images. The core of this study involves the

extraction of image attributes from a pretrained convolutional neural network (CNN), which subsequently enhances the Bayesian support vector machine classification process. To ascertain the most effective network for cancer detection, pretrained neural networks such as AlexNet, VGG-16, and Inception-V3 were evaluated. Furthermore, the effectiveness of this proposed approach was examined across four distinct databases. Among these, two databases were sourced from Indian hospitals, varying in magnifications (4X, 10X, 20X, and 40X), while the remaining two were gathered from diverse photographic platforms. Through a comparison of the approach with publicly available data, it was established that the Inception-V3 platform achieved an accuracy range of 96.5% to 99% in testing, aligning closely with the performance of the provided method. This highlights the potency of the proposed methodology in pancreatic cancer detection.

Another study conducted by Rana, M. and Bhushan, M. in the year 2023 [25], this paper describes how to learn a migration model without using additional methods and post-processing to change tumor color. For BreakHis datasets swelling order, another component called relative accuracy is well estimated from the raw data (87%) for the transfer class (LENET, DarkNet53, VGG16, ResNet50, DarkNet19, Inception, and Xception).

According to Shallu Sharma and Sumit Kumar in their study conducted in the year 2022 [26], a trained exception model demonstrated its ability to classify breast cancer images based on manual enhancement methods. The Xception model with SVM key and 40X “Radiation Basis” kernel at 100X, 200X and 400Ks magnifications had the best performance of 96.25%, 96.25%, 96.25% and 94.11% respectively.

Another study conducted by R. Man, P. Yang and B. Xu in the year 2020 [27], a novel approach titled "Densnet121-anogan" has been introduced to categorize histopathology images into benign and malignant classifications. This innovative method is structured around two key components: an unsupervised anomaly detection process employing an adaptive adversarial network (AnoGAN) for identifying contaminated patches, and a convolutional network (DenseNet) responsible for extracting multilayer features from distinguishing patches. The efficacy of this proposed technique was thoroughly assessed through a 5-fold crossover examination on the publicly accessible BreakHis database. Impressively, the results demonstrated exceptional performance with a remarkable accuracy of 99.13% at the image level, particularly notable for images captured at 40X magnification. This underscores the potential and promise of the "Densnet121-anogan" approach in the accurate classification of histopathology images in the context of cancer diagnosis.

A study by A. S. Alhanaf, S. Al–jumaili, G. Bilgin, A. D. Duru, S. Alyassri and H. H. Balik in the year 2022 [28], a rapid, accurate, appropriate and reliable rectal cancer classification system is targeted and established. Convolutional neural networks (CNNs) can be harnessed to analyze and classify digital histology images through pixel count-based methodologies, furnishing extensive insights into colon cancer. They used two types of classification methods (simple classification and K-fold cross-validation) and ten pre-trained models to classify tumor tissue (low hump, high hump). Among these eight deep transfer learning models, DarkNet53 showed the highest accuracy - 96.6% or 5 times higher, while ResNet50 showed the highest result for a regular classifier - 98.7%.

Another study by Sethy, P.K. and Behera, S.K. in the year 2022 [29], a classification algorithm based on artificial neural networks (ANN) has been developed, aiming to diagnose

breast cancer by amalgamating comprehensive manual features with detailed analysis of histological images. The system discerns four distinct levels of damage: total, severe, environmental, and invasive. This classification is orchestrated through three overarching categories, each composed of three broad and three specific attributes. To enhance the classification process, the potency of deep learning is harnessed through the FC6 layer of three trained networks: AlexNet, VGG16, and VGG19. After establishing guideline scores for GLCM, HOG, and LBP, the most effective classifier is determined, considering both depth features and manually generated attributes. Employing a clustering model grounded in the K-nearest neighbors (KNN) algorithm, a fusion of VGG16 and LBP features yields a notable detection efficiency of 84.2%. This system showcases the potential of integrating deep learning and manual features for the accurate identification of various levels of breast cancer damage.

Another study by Mohammed Senan, Ebrahim and some other authors in the year 2021 [30], a technique employing a neural network, specifically AlexNet, was formulated to extract prevalent features from the BreakHis database with the aim of distinguishing between benign and malignant tumors. To examine this hypothesis, researchers conducted four separate experiments, each corresponding to varying magnifications (40X, 100X, 200X, and 400X). Within each experiment, a total of 1407 images were included. The network underwent training and testing using an 80% - 20% split for training and testing images, respectively. As a result of this approach, the devised system achieved an impressive accuracy of 95%.

III. SIGNIFICANT CHALLENGES AND MOTIVATION

There are many problems with the work of the brain cancer detection. Due to which the efficiency of the machine decreases. Classification usually involves two steps – separating the shape and grouping it. Feature extraction is an important part of pattern recognition [31]. The objects are predicted based on important features such as size, color, name, etc. However, the efficiency of the pool depends on the power band. The recent achievements of deep learning within the medical domain have captivated the attention of computer vision scholars. Nevertheless, it's important to acknowledge that not all deep learning models can be readily tailored for accurate classification and application, often due to extended training durations. In the present work, transfer learning is being implemented to get the desired result with various models.

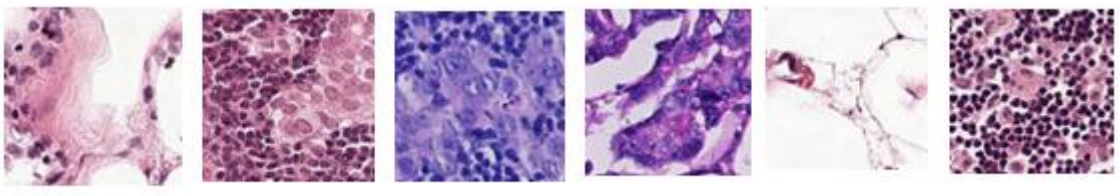


Figure 1: Histopathological images of Brain Cancer.

In this study, we present a study to distinguish two types of brain cancer. An additional correction technique was applied to solve manual measurement problems, which was complemented by histogram comparison. When components are downloaded from two CNN models [18,24,31], transfer learning is performed and integration is performed. The

combination of two types of CNNs is to obtain a new image vector with the data. Although this approach improves accuracy, computation time suffers. We have proposed a selection strategy to increase the processing and computation time. The complex benefits obtained with this approach are shared with the model we wish to implement. Existing binary cancer classification is not a new task [32,33,34], but using histopathological images this same task becomes a bit tedious and non-trivial as it contains some values which is not continuous. Histopathological images are nothing but tissue images that have been magnified 1000 times to get a clear view. So, simply we are extrapolating the original images i.e. when we magnify a image to 1000x times, there are high chances of pseudo values interfering in it. When the tissue was captured, it had some natural colors, but as we magnify it to a value of 1000x, so the colour values also changes and pseudo colors appear. This becomes difficult for the machine to learn such pseudo colors and so this process becomes non-trivial. So there is high chance to get pseudo colors from the images which subsequently lead to erroneous result. From feature based map and our proposed model, we have tried to reduce the computational time as well as the amount of training data. The motivation behind this project is to get an higher accuracy by utilizing less amount of data and by minimizing the computational complexity.

IV. PROPOSED METHODOLOGY

1. **Image Pre-Processing**: A substantial portion of the pixels present in an image are superfluous and do not substantially contribute to the overall image content. When engaging with AI networks, these redundant pixels should be deactivated to prevent undue computational burden. This task can be accomplished through compression techniques. The process commences with initiating an image editor within the database, often facilitated using the OpenCV library in the Python programming language. Multiple alternative modules are available for this process, which can be sourced from various libraries or other dedicated image processing software. This is necessary to eliminate input redundancy, which increases the computational complexity of the network, but does not provide a significant improvement in the output. Both dimensions are reduced to 2, giving a 96 x 96 pixel image, thus preserving the slide's original aspect ratio. In our project, the dataset have been collected from kaggle [35].

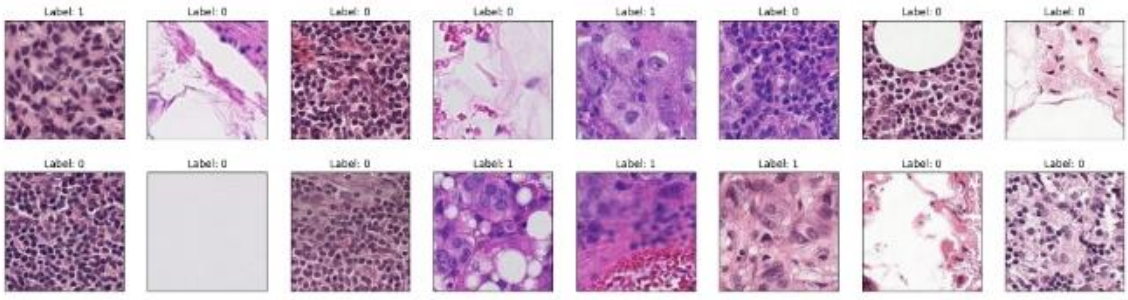


Figure 2: Histopathological images from dataset

2. **Feature extraction:** Feature training is an important part of the classification of human and machine algorithms. According to a study, human brain is more sensitive to shapes while computers are more sensitive to shapes and patterns. For this reason, active learning is different from manual learning than machine learning. On the surface, malignant carcinomas have large irregular nuclei or multiple nuclear arrangements. There are also changes in the cytoplasm where new substances appear or normal substances disappear. Damaged cells have very little cytoplasm and are often vacuolated. All these are analyzed by experts or algorithms are developed to measure these objects for automatic detection. This approach is complex and imperfect because there are unknown selection and estimation errors that are difficult to handle. In the case of targeted training, we don't need to explicitly offer this option. In this case, images are displayed in a CNN-like format that includes results (positive or negative). Activating the filtering system during the training process allows the CNN to extract computational resources. Briefly, the CNN filter of a building, its weights and its materials are used in the model estimation process. Thus, CNN takes pixels from the image and sends them through a scientifically weighted filtering process. These weights are fed into the deep tissue complex architecture for final prediction. The CNN models used in this comparison are VGG19, ResNet50, InceptionV3, and Gabor Filter.

- **Gabor Filter:** The Gabor filter serves as a linear filter applied in tissue analysis, primarily focusing on the examination of frequency content at specific points or orientations within the analyzed region.

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right)$$

Where, $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$.

In the equation, λ denotes the wavelength of the sine factor, θ denotes the normal orientation of the band parallel to the Gabor function, Ψ is the phase shift, σ is the sigma/standard deviation of the Gaussian envelope and γ is the aspect ratio. spatially, and denotes the ellipticity of the carrier of the Gabor function.

- **VGG19 Model Architecture:** VGG19 belongs to the VGG model family and is comprised of a total of 19 layers, encompassing 16 convolution layers, 3 fully connected layers, 5 maxpool layers, and a single softmax layer. Within the VGG model family, there exist additional variations such as VGG11, VGG16, and more.

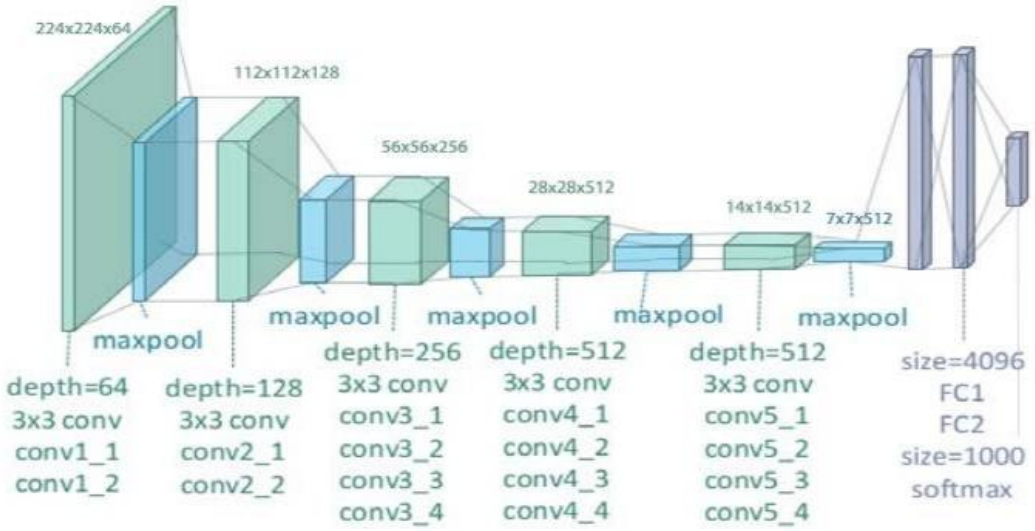


Figure 3: VGG19 Model Architecture [36]

- **Inception V3 Model Architecture:** Inception-v3 is a convolutional neural network characterized by 48 depth levels. It is feasible to employ a pre-existing iteration of this network, which has been trained on an excess of 1 million images derived from the ImageNet database [38]. This pre-trained model possesses the ability to categorize images into 1000 distinctive classes of features. Consequently, the network acquires intricate and comprehensive representations of features across a vast array of images.

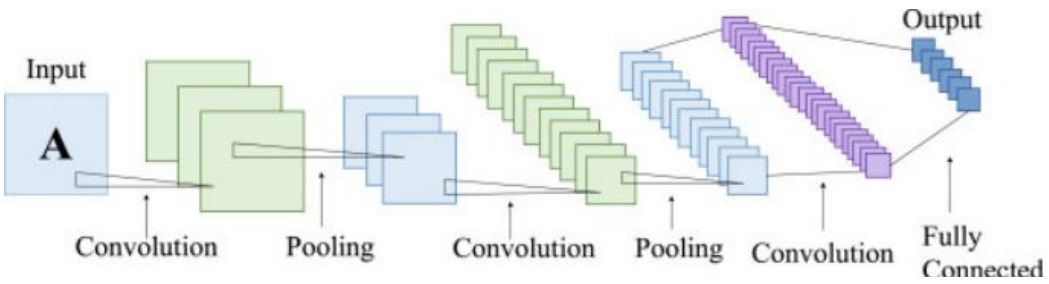


Figure 4: Inception V3 Model Architecture [37]

- **ResNet50 Model Architecture:**

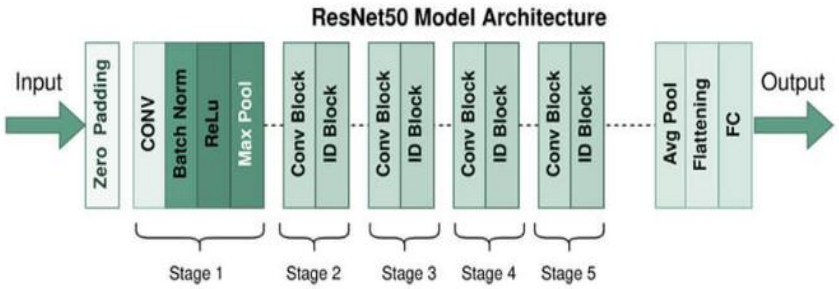


Figure 5: ResNet50 Model Architecture [39]

ResNet50 stands as a version of the ResNet model, encompassing 48 convolutional layers, along with a single maxpool and an average pool layer. The ResNet architecture is extensively employed, and our investigation particularly delved into the intricacies of the ResNet50 structure.

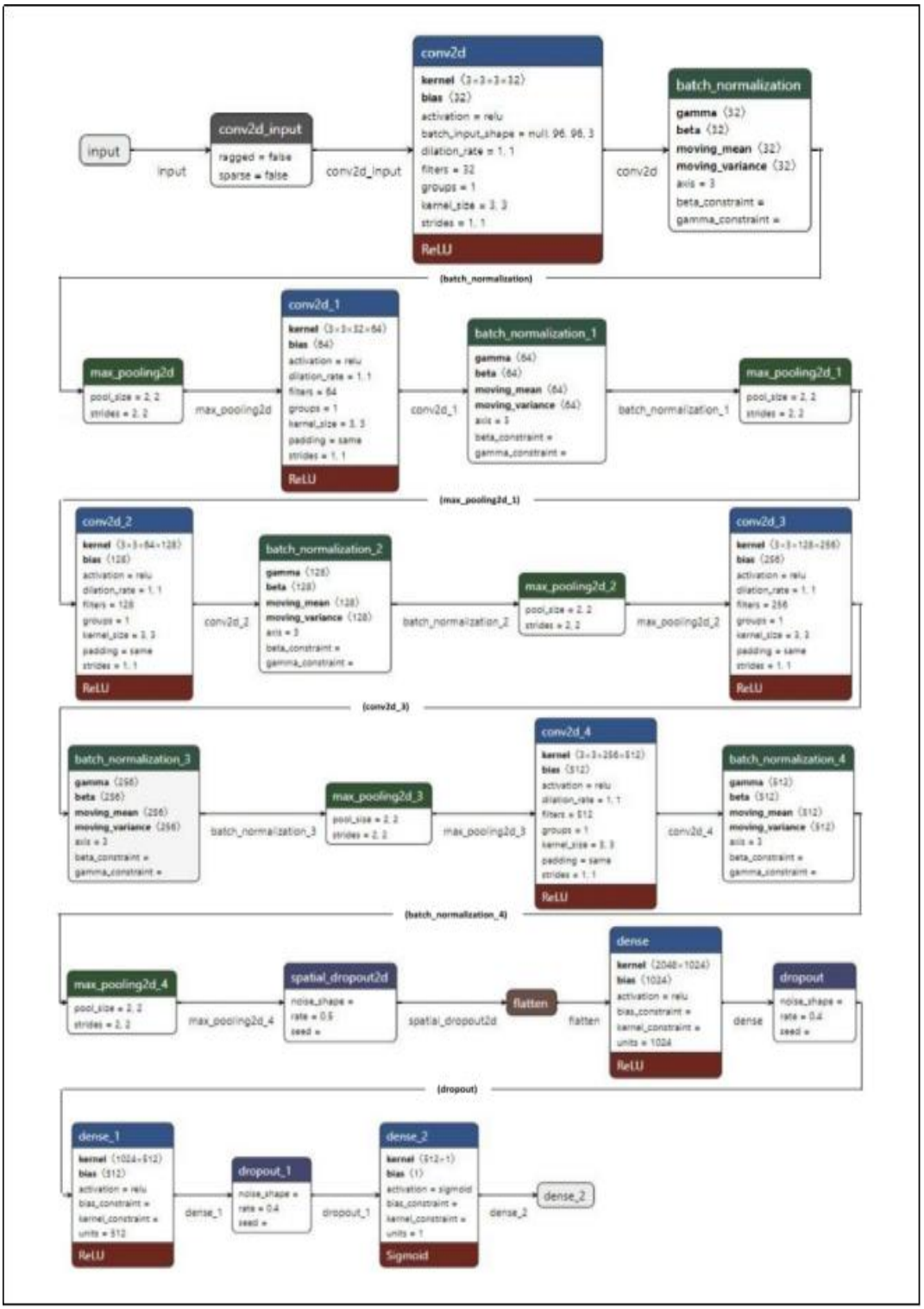


Figure 6: Our proposed model

3. **Classification:** The clustering procedure involves utilizing the weight map derived from the concluding clusters, which is subsequently fed into a fully connected network. This network computes the loss and adjusts the values of the internal hidden nodes accordingly.

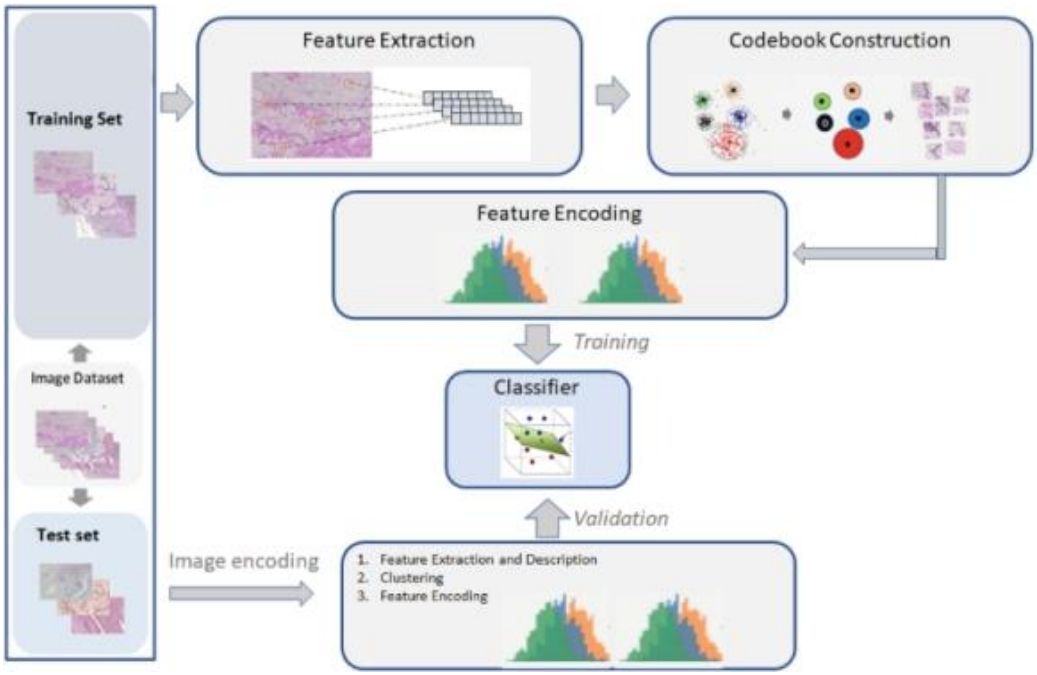


Figure 7: Classification of an image

4. Objectives of the proposed project

- Designing a robust feature extractor.
- Designing & implementing a classifier.
- Study performance of proposed classifier against state of art.
- To extend the work for detection of other diseases.

V. RESULT AND DISCUSSION

For the analysis and comparison of our model with the pre-trained models that have been developed earlier, we have used various metrics. Following are the guide to the metrics:

1. **Accuracy:** Accuracy serves as a metric that quantifies a model's performance across various classes. It's particularly useful when all classes are of equal importance. This metric is computed by dividing the number of accurate predictions by the total number of predictions made.

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

2. **Precision:** The precision score is determined by the proportion of positive samples over the total number of samples classified as positive (whether truly positive or negative). Accuracy, on the other hand, gauges how well the model performs when identifying positive instances.

$$\Pr ecision = \frac{True_{positive}}{True_{positive} + False_{positive}}$$

3. **Recall:** Recall represents the proportion of correctly identified good examples compared to the total number of original good examples that were classified correctly. It's important to note that recall measures the model's ability to find accurate patterns. The larger the memory capacity, the higher the quality of the model's performance.

$$\mathrm{Re}\, call = \frac{True_{positive}}{True_{positive} + False_{negative}}$$

4. **ROC-AUC Curve:** ROC or receiver operating curve is a probability plot that shows how a class performs at different levels and draw a line between the two dimensions, which are:
   - TPR or True Positive Rate
   - FPR or False Positive Rate

   In the curve, FPR is on the X-axis and TPR is plotted on Y-axis .

   - **TPR (True Positive Rate):**
     TPR or True Positive rate is a synonym for Recall, which can be calculated as:

$$TPR = \frac{TP}{TP + FN}$$

   - **FPR or False Positive Rate :**
     It can be calculated as

$$FPR = \frac{FP}{FP + TN}$$

     Here, TP denotes True Positive
     FP denotes False Positive
     TN denotes True Negative
     FN denotes False Negative

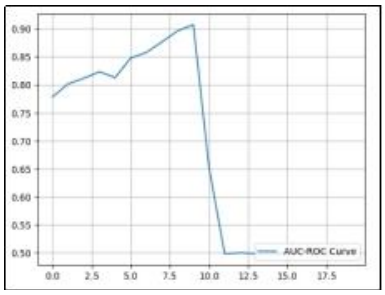
     Now, to efficiently calculate the values at any threshold level, we need a method, which is AUC.

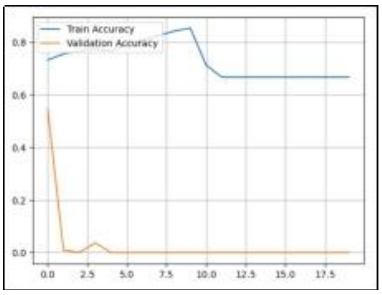
5. **AUC (Area Under the ROC Curve):** As its name suggests, AUC (Area Under the Curve) is calculated by measuring the two-dimensional area under the ROC (Receiver Operating Characteristic) curve. The ROC curve provides a graphical representation of a binary filter's performance at different thresholds, while the AUC serves as a quantitative summary measure of this performance. It calculates the area from (0,0) to (1,1) across the

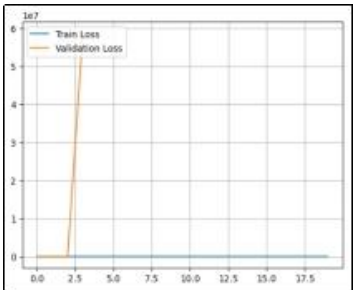
entire ROC curve. AUC values range between 0 and 1, with a higher value indicating better discrimination and a model that performs well. Ideally, a good model will have an AUC close to 1, highlighting its ability to accurately distinguish between classes**.**

Now the result obtained from various models are given below:

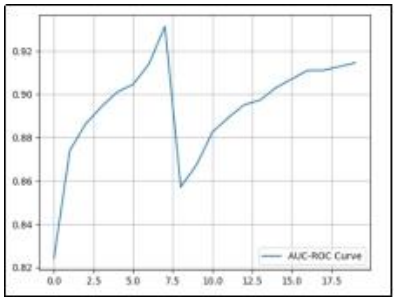
6. VGG19:

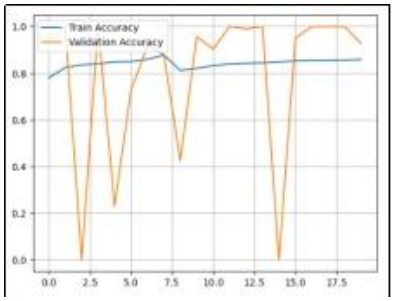


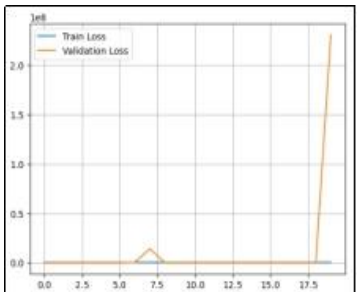




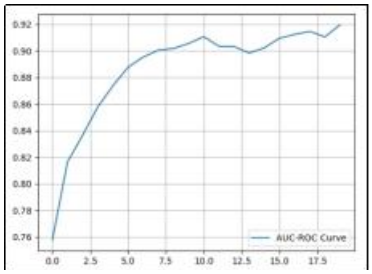
For sample size = 32

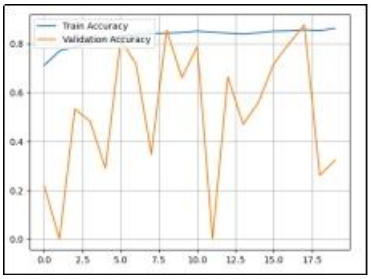


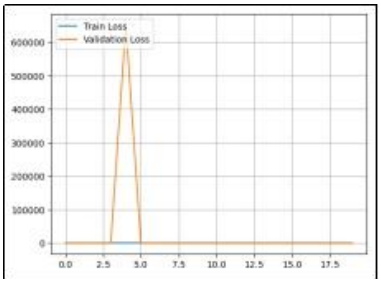




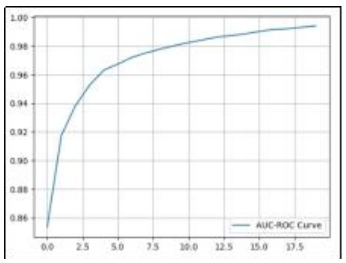
For sample size = 64

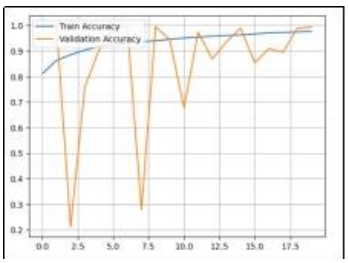


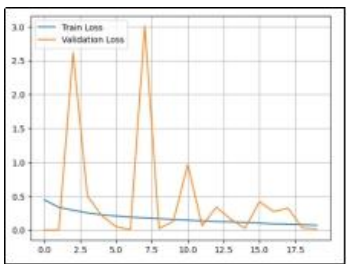




For sample size = 128

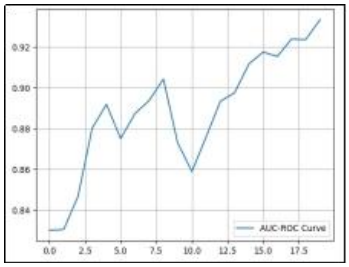


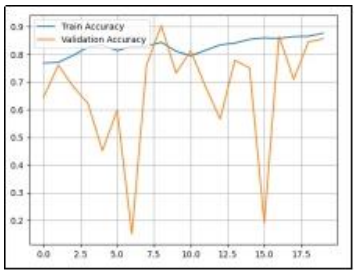


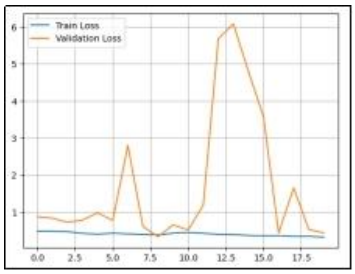


For sample size = 256

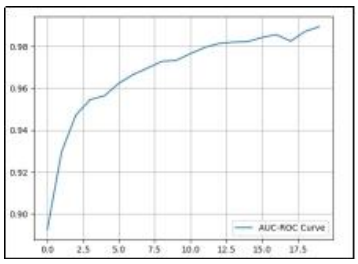
7. InceptionV3:

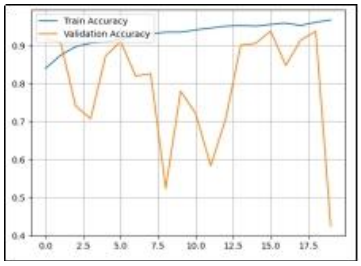


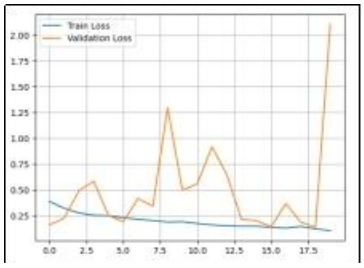




For sample size = 32

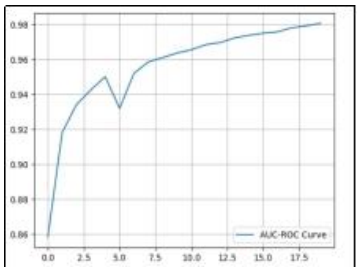


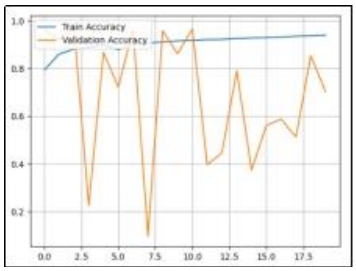


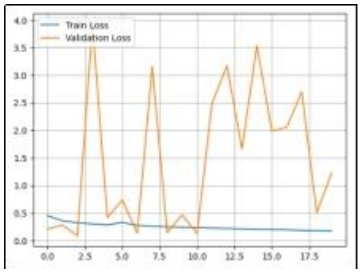


For sample size = 64

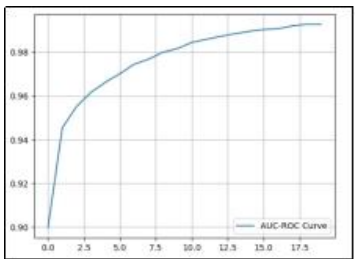
8. Resnet50:

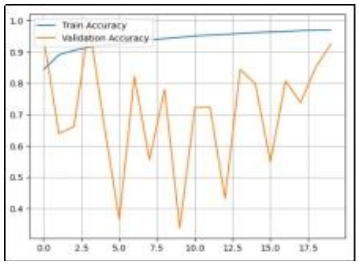


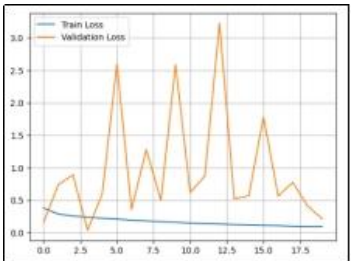




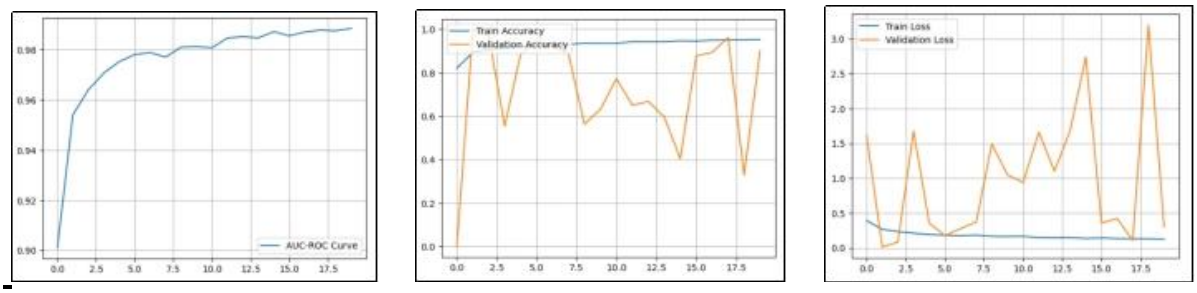
For sample size = 32







For sample size = 64



For sample size = 256

9. 1.7.9 Gabor Filter:

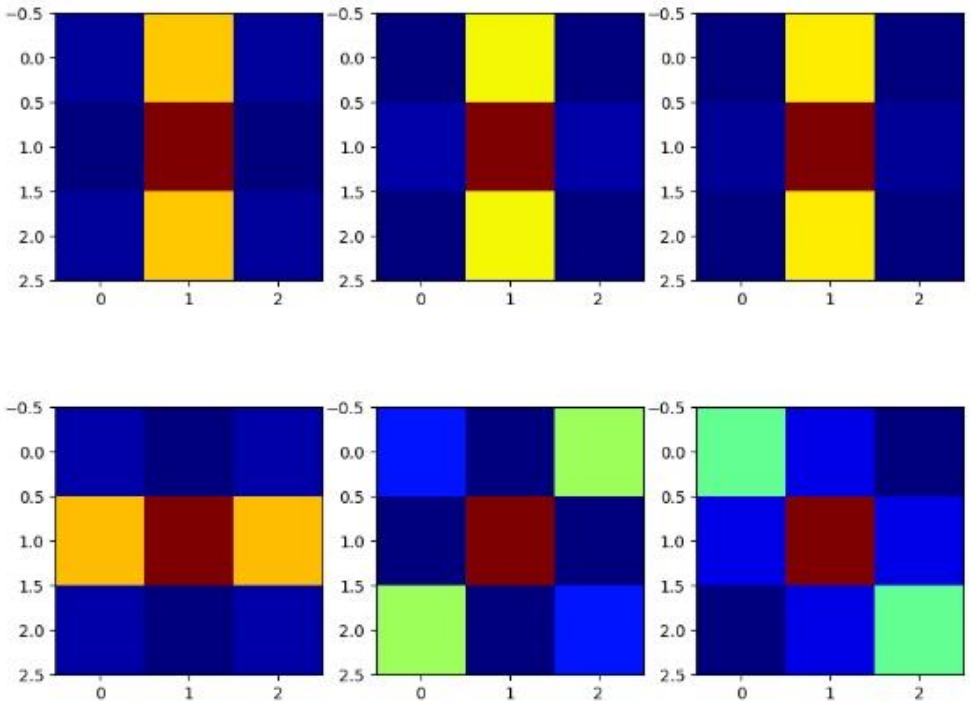
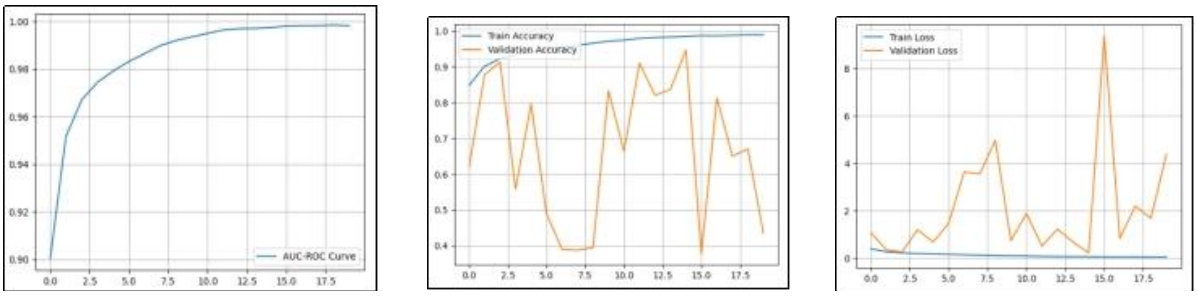


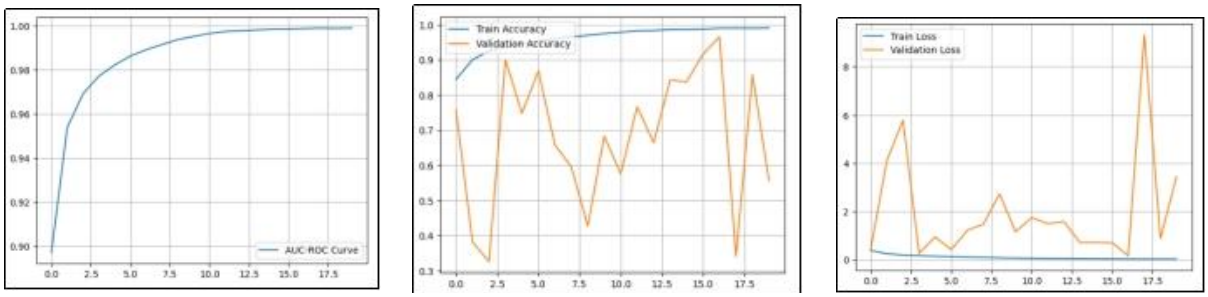
Figure 8: Gabor kernels for different values of lamda (λ) and psi (Ψ)

Figure 9: Result of gabor feature based classifier

10. Our Proposed Model:



For sample size = 128



For sample size = 256

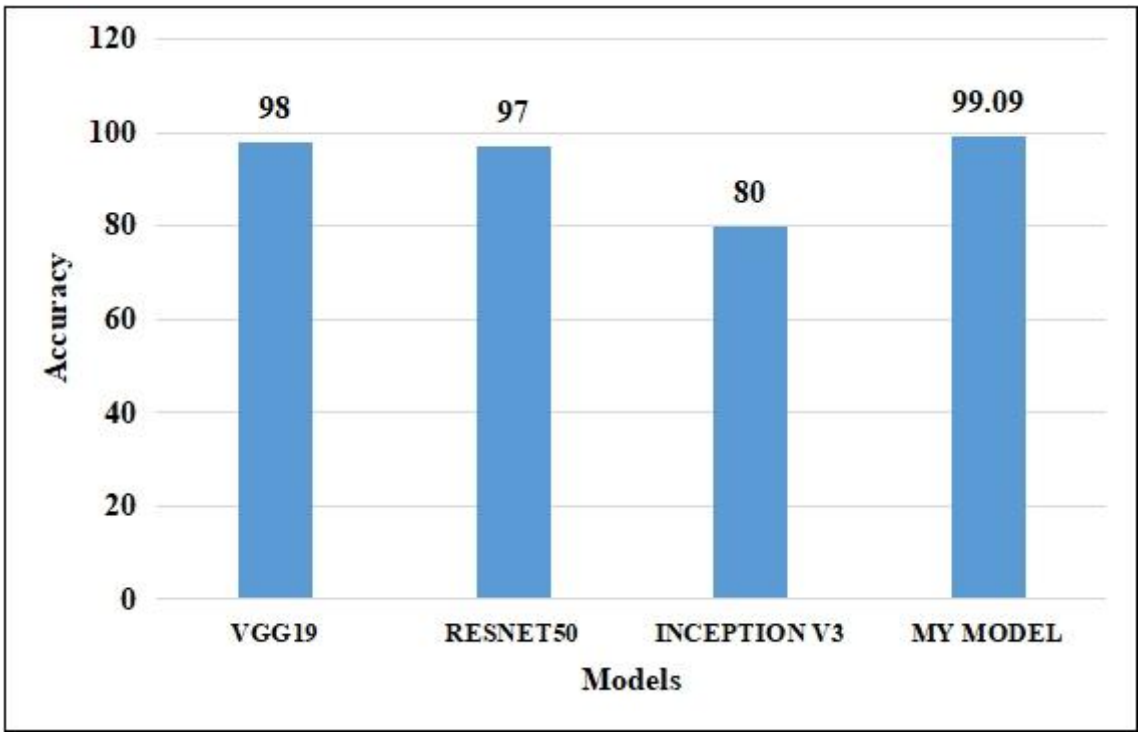
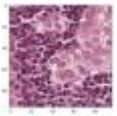


Figure 10: Accuracy of various models compared

A test has been conducted by taking one sample image and passing through our designed model.

Test image	Actual value	Predicted value	Image size
	0	0.00188882 ≈ 0	(1,96,96,3)

From the above data, it is clear that our model has worked very well giving accuracy of 99.09% and AUC of 99.89%. All the batch sizes with respective models have been tested for 20 epochs. Therefore, we can conclude that the model designed by us is working well against other models that have been designed earlier.

VI. CONCLUSION AND FUTURE SCOPE

A good model has been developed for the identification of histopathological images of cancer with correct classification. We have developed a good model [Figure-6] to increase the durability of the product. The model utilizes a dataset that is divided such that 75% of the images are used for training, 15% for testing, and 10% for validation purposes. In fact, the action plan provides a model for processing/correction - ID the bad distribution of the image from the input stage to the output stage. The effectiveness of the most advanced approach is validated by achieving a remarkable 99% accuracy through the optimization of specific parameters. Furthermore, a comparative analysis between the CNN model and existing methodologies demonstrates the superior accuracy of the proposed model. In this context, we are delighted to present an automated binary classification method that leverages the power of deep learning for precise brain tumor classification. A deep neural network model was developed and used as a histopathological image retrieval tool to identify patterns in histopathological images. All layers have been modified and all deep neural networks have been trained on classes to differentiate between these two types of cancer. The proposed model has about 4.2 million training parameters, compared to 20.1 million, 23.9 million, and 21.8 million for VGG19, ResNet50, and InceptionV3, respectively. In conclusion, as far as our knowledge is concerned, the reported results are better than cancer detection.

Cancer diagnosis is based on histological features. With advances in machine learning and image processing algorithms, the ability to detect complex patterns in samples that represent cancer progression is increasing. In the future, we will see advances in imaging technology that can differentiate between different types of cancer and determine the exact stage of cancer. Advances in diagnosis and prognosis can also be achieved, saving time and improving patient outcomes.

REFERENCES

[1] https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html
[2] https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21654
[3] https://www.who.int/news-room/fact-sheets/detail/cancer
[4] https://dmimed.com/
[5] https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri
[6] https://paperswithcode.com/task/histopathological-image-classification
[7] Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. IEEE Rev Biomed Eng. 2009;2:147-71. doi: 10.1109/RBME.2009.2034865. Epub 2009 Oct 30. PMID: 20671804; PMCID: PMC2910932.

[8] Rosenzweig MQ, Gardner D, Griffith B. The History and Physical in Cancer Care: A Primer for the Oncology Advanced Practitioner. J Adv Pract Oncol. 2014 Jul-Aug;5(4):262-8. doi: 10.6004/jadpro.2014.5.4.3. PMID: 26110070; PMCID: PMC4457181.

[9] Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Humaniz Comput. 2022 Jan 13:1-28. doi: 10.1007/s12652-021-03612-z. Epub ahead of print. PMID: 35039756; PMCID: PMC8754556.

[10] https://www.mvtec.com/technologies/deep-learning/classic-machine-vision-vs-deep-learning

[11] Walsh, Joseph & O' Mahony, Niall & Campbell, Sean & Carvalho, Anderson & Krpalkova, Lenka & Velasco-Hernandez, Gustavo & Harapanahalli, Suman & Riordan, Daniel. (2019). Deep Learning vs. Traditional Computer Vision. 10.1007/978-3-030-17795-9_10.

[12] Pralhad Gavali, J. Saira Banu, Chapter 6 - Deep Convolutional Neural Network for Image Classification on CUDA Platform, Editor(s): Arun Kumar Sangaiah, Deep Learning and Parallel Computing Environment for Bioengineering Systems,Academic Press,2019,ISBN 9780128167182, https://doi.org/10.1016/B978-0-12-816718-2.00013-0.

[13] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, Yujun Liao,Review the state-of-the-art technologies of semantic segmentation based on deep learning,Neurocomputing,Volume 493,2022,, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2022.01.005.

[14] Z. Xiao and L. N. Feng, "A Study on Wireless Capsule Endoscopy for Small Intestinal Lesions Detection Based on Deep Learning Target Detection," in IEEE Access, vol. 8, pp. 159017-159026, 2020, doi: 10.1109/ACCESS.2020.3019888.

[15] Liu, Z., Tong, L., Chen, L. et al. Deep learning based brain tumor segmentation: a survey. Complex Intell. Syst. 9, 1001–1026 (2023). https://doi.org/10.1007/s40747-022-00815-5

[16] Yashwant Kurmi, Vijayshri Chaurasia, Narayanan Ganesh,Tumor Malignancy Detection Using Histopathology Imaging,Journal of Medical Imaging and Radiation Sciences,Volume 50, Issue 4,2019, ISSN 1939-8654, https://doi.org/10.1016/j.jmir.2019.07.004.

[17] Khan MA, Ashraf I, Alhaisoni M, Damaševičius R, Scherer R, Rehman A, Bukhari SAC. Multimodal Brain cancer Classification Using Deep Learning and Robust Feature Selection: A Machine Learning Application for Radiologists. Diagnostics (Basel). 2020 Aug 6;10(8):565. doi: 10.3390/diagnostics10080565. PMID: 32781795; PMCID: PMC7459797.

[18] Q. A. Al-Haija and A. Adebanjo, "Breast Cancer Diagnosis in Histopathological Images Using ResNet-50 Convolutional Neural Network," 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Vancouver, BC, Canada, 2020, pp. 1-7, doi: 10.1109/IEMTRONICS51293.2020.9216455.

[19] Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R. A., Rehman, A., Iqbal, M., & Saba, T. (2021). Skin cancer detection from dermoscopic images using deep learning and fuzzy k -means clustering. Microscopy Research and Technique. doi:10.1002/jemt.23908

[20] Naresh Kumar, Manoj Sharma, Vijay Pal Singh, Charanjeet Madan, Seema Mehandia, An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images, Biomedical Signal Processing and Control,Volume 75,2022,103596, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2022.103596.

[21] Srikantamurthy, M.M., Rallabandi, V.P.S., Dudekula, D.B. et al. Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. BMC Med Imaging 23, 19 (2023). https://doi.org/10.1186/s12880-023-00964-0

[22] Mangal Sanidhya , Chaurasia Aanchal , Khajanchi Ayush , Convolution Neural Networks for diagnosing colon and lung cancer histopathological images, 2020, https://doi.org/10.48550/arXiv.2009.03878.

[23] Sumaiya Dabeer, Maha Mohammed Khan, Saiful Islam, Cancer diagnosis in histopathological image: CNN based approach, Informatics in Medicine Unlocked,2019,100231,ISSN 2352-9148, https://doi.org/10.1016/j.imu.2019.100231.

[24] Babu, Tina et al. ‘Colon Cancer Prediction on Histological Images Using Deep Learning Features and Bayesian Optimized SVM’. 1 Jan. 2021 : 5275 – 5286.

[25] Rana, M., Bhushan, M. Classifying breast cancer using transfer learning models based on histopathological images. Neural Comput & Applic (2023). https://doi.org/10.1007/s00521-023-08484-2.

[26] Shallu Sharma, Sumit Kumar,The Xception model: A potential feature extractor in breast cancer histology images classification,ICT Express,Volume 8, Issue 1,2022,ISSN 2405-9595, https://doi.org/10.1016/j.icte.2021.11.010.

[27] Man, P. Yang and B. Xu, "Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screened by Generative Adversarial Networks," in IEEE Access, vol. 8, pp. 155362-155377, 2020, doi: 10.1109/ACCESS.2020.3019327.
[28] S. Alhanaf, S. Al–jumaili, G. Bilgin, A. D. Duru, S. Alyassri and H. H. Balik, "Deep Transfer Learning Methods for Classification Colorectal Cancer Based on Histology Images," 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 2022, pp. 818-821, doi: 10.1109/ISMSIT56059.2022.9932746.
[29] Sethy, P.K., Behera, S.K. Automatic classification with concatenation of deep and handcrafted features of histological images for breast carcinoma diagnosis. Multimed Tools Appl 81, 9631–9643 (2022).
[30] Mohammed Senan, Ebrahim, Fawaz Waselallah Alsaade, Mohammed Ibrahim Ahmed Al-mashhadani, Theyazn H. H aldhyani, and Mosleh Hmoud Al-Adhaileh. 2021. "Classification of Histopathological Images for Early Detection of Breast Cancer Using Deep Learning." Journal of Applied Science and Engineering 24: 323–29, https://doi.org/10.6180/jase.202106_24(3).0007.
[31] Rajinikanth, V.; Joseph Raj, A.N.; Thanaraj, K.P.; Naik, G.R. A Customized VGG19 Network with Concatenation of Deep and Handcrafted Features for Brain Tumor Detection. Appl. Sci. 2020, 10, 3429. https://doi.org/10.3390/app10103429.
[32] Ding, S., Zhu, H., Jia, W. et al. A survey on feature extraction for pattern recognition. Artif Intell Rev 37, 169–180 (2012). https://doi.org/10.1007/s10462-011-9225
[33] T. T. L. Vuong, B. Song, K. Kim, Y. M. Cho and J. T. Kwak, "Multi-Scale Binary Pattern Encoding Network for Cancer Classification in Pathology Images," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 3, pp. 1152-1163, March 2022, doi: 10.1109/JBHI.2021.3099817.
[34] Han, X., Li, D., Liu, P. et al. Feature selection by recursive binary gravitational search algorithm optimization for cancer classification. Soft Comput 24, 4407–4425 (2020). https://doi.org/10.1007/s00500-019-04203-z
[35] M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning," 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, Turkey, 2018, pp. 1-4, doi: 10.1109/EBBT.2018.8391453.
[36] https://www.kaggle.com/competitions/histopathologic-cancer-detection/data
[37] Zheng, Yufeng & Yang, Clifford & Merkulov, Aleksey. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. 4. 10.1117/12.2304564.
[38] Dong, N., Zhao, L., Wu, C. H., & Chang, J. F. (2020). Inception v3 based cervical cell classification combined with artificially extracted features. Applied Soft Computing, 93, 106311. doi:10.1016/j.asoc.2020.106311
[39] ImageNet. http://www.image-net.org
[40] https://www.educba.com/keras-resnet50/

