Summer 8-9-2022

Deep Learning for Classification of Brain Tumor Histopathological Images

Ifeanyi Austin Ezuma
Georgia State University

Deep Learning for Classification of Brain Tumor Histopathological Images

by

Ifeanyi A. Ezuma

Under the Direction of Yi Jiang, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2022

ABSTRACT

Histopathological image classification has been at the forefront of medical research. We evaluated several deep and non-deep learning models for brain tumor histopathological image classification. The challenges were characterized by an insufficient amount of training data and identical glioma features. We employed transfer learning to tackle these challenges. We also employed some state-of-the-art non-deep learning classifiers on histogram of gradient features extracted from our images, as well as features extracted using CNN activations. Data augmentation was utilized in our study. We obtained an 82% accuracy with DenseNet-201 as our best for the deep learning models and an 83.8% accuracy with ANN for the non-deep learning classifiers. The average of the diagonals of the confusion matrices for each model was calculated as their accuracy. The performance metrics criteria in this study are our model's precision in classifying each class and their average classification accuracy. Our result emphasizes the significance of deep learning as an invaluable tool for histopathological image studies.

INDEX WORDS: Histopathological images, Deep learning model, Non-deep learning model, Brain tumor, Machine learning classifier, Histogram of gradient

Deep Learning for Classification of Brain Tumor Histopathological Images

by

Ifeanyi A. Ezuma

Committee Chair: Yi Jiang

Committee: Gengsheng Qin

Yichuan Zhao

Jun Kong

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2022

DEDICATION

To the Almighty God for taking me this far in life. Also, I dedicate this work to the memory of my late father, Louis Ezuma, and my amazing mother, Theresa Ezuma. Their unwavering love and sacrifice paved the way for my success throughout the years.

To my loving family, who has always supported me. To my brother Dr. Martins Ezuma for his relentless advice and support, Ijeoma Nriagu, Chidinma Eze, Cynthia Ezuma, and Nnenna Ezuma. Their love knows no bounds. I would also like to thank my nephews and nieces; Osinachi Ezuma, Ebuka Ezuma, Chinemerem Eze, Somtochukwu Eze, Chimamanda Eze and Chizara Nriagu.

To my wonderful friends, Cindy Okoh, Seyifumi Owoleye, Adeniyi Abolade, Olakunle Sodiya, Zviko Machikiti, Foyin Bamidele, Oluwatola Femi-Olatunji, Yetunde Awodeko, Chuka Nwalibe, and Samuel Osuji.

To Dr. Olusiji Medaiyese and Dr. Ugochukwu Nsofor for their comradeship during my master's program.

ACKNOWLEDGEMENTS

I would like to thank God for His grace, favor, and strength to undertake this journey. I would also like to thank my academic supervisor and mentor Dr. Yi Jiang for her invaluable guidance and supervision. I will be forever grateful for her contributions to my academic development. Lastly, I would like to thank the members of my master's thesis committee, who helped model me as an academic student and helped shape my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- Whole slide imaging (WSI)
- Machine learning (ML)
- Artificial intelligence (AI)
- Deep learning (DL)
- Non-deep learning (non-DL)
- Regio of interest (ROI)
- Convoluted neural network (CNN)
- Real non-linear function (ReLU)
- Local binary pattern (LBP)
- Scale-invariant future transformation (SIFT)
- Histogram of oriented gradient (HOG)
- Isocitrate dehydrogenase 1 (IDHI)
- Support Vector Machine (SVM)
- Naïve Bayes (NB)
- Decision tree (DT)
- K-nearest neighbor (KNN)
- Computer-aided design (CAD)
- Gray level co-occurrence matrix (GLCM)
- Quadratic Discriminant analysis (QDA)
- Random forest (RF)
- Hematoxylin & Eosin (H&E)
- Radial Basis Function (RBF)

- Linear discriminant analysis (LDA)
- Area under the curve (AUC)
- Support vector regression (SVR)
- Bidirectional long-short memory (BLSTM)
- Spatially constrained CNN (SC-CNN)
- Speeded-up robust features (SURF)
- Astrocytoma grade II (AII)
- Astrocytoma grade III (AIII)
- Oligoastrocytoma grade II (OAII)
- Oligoastrocytoma grade III (OAIII)
- Oligodendroglioma grade II (OII)
- Oligodendroglioma grade III (OIII)
- Glioblastoma Multiforme (GBM)
- Secondary glioblastoma multiforme (GBMII)
- Artificial neural network (ANN)

1 INTRODUCTION

Digital pathology represents an essential evolution in modern medicine. Although there were increasing advancements from as far back as the 17th century, image analysis was not making much progress until the advent of digital imaging and computational research in the second half of the last century. Digital pathology is referred to as the practice of pathology using digital imaging. Recently, its emergence has become a vital tool for making cancer prognosis and diagnosis [1]. It is the practice of pathology using digital imaging. Histopathology is a clinical medical procedure involving examining tissue removed from the patient for a comprehensive study. Histopathological images are very influential in deciding the final approach when determining effective treatments; they are significant when investigating a particular biological structure and diagnosing several diseases [2].

Furthermore, the deductions from a histopathology image remain the gold standard in diagnosing almost all significant types of cancer. The usage of digital pathology has employed live streaming of images, static images, and whole slide imaging (WSI). The process of WSI involves digitizing glass slides of collected samples of cells or tissues and then staining them [3]. The resultant images are called histopathological images. However, its evolution has become hampered by the lack of available resources, large dimensionality of WSI images, technology limitations, stain variability across laboratories, insufficient training samples, cost of equipment, etc. These images are huge, have high-level, complex clinical features, and only represent a few annotated regions [1]. Nevertheless, it is significant for the future practice of pathology [4].

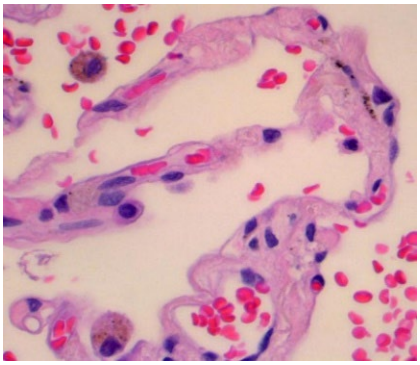


Figure 1.1: An example of a histopathology image.

The individuality of histopathological images has stimulated efforts to establish novel automated image analysis methodologies [1]. In recent years, the advancement of computational techniques has transformed the practice of pathology. They can be applied to histopathological images to specify regions of interest, make a diagnosis and extract features that may relate to treatment and prognosis. The proliferation of machine learning (ML) algorithms has facilitated significant support for medical research and clinical studies [3]. Artificial Intelligence (AI) models have increasingly migrated from conventional ML to deep learning (DL) because of their ability to automatically learn features from data in a more accessible, precise, and accurate manner [1]. Thus, the rise of powerful computational resources has diverted the interest of DL models in a broad range of medical image applications, such as interpreting, analyzing, and extracting pertinent information from WSI [1].

1.1 Deep Learning

Until recently, most techniques employed for medical image analysis relied on traditional non-deep learning (non-DL) models. Objects of interest in histopathological images are often too complex to be represented explicitly by any simple mathematical equation or model [5]. It usually requires a compound model with many parameters that cannot be achieved manually and

is dependent on data. Thus, the role of non-DL models has been significant in medical imaging. The schematic of such models mostly starts with data preprocessing and preparation, then regions of interest (ROI) selection, feature extraction, feature selection, and eventually, classification using linear and non-linear models [6]. The objective of these non-DL models is to determine optimal discrimination between the multiple output classes vis training. However, the performances of these ML models mainly rely on the selection of features on which they are being trained, despite having made satisfactory progress in analyzing medical images [7].

In contrast to non-DL approaches, DL models differ in their architectural details. These models consist of an arbitrary number of layers of list-based vectors, or neurons which connect inputs to subsequent layers sequentially through weights vectors. Each layer can be used as a threshold by an activation function to transform further the output passed on to the succeeding layers. These intricate non-linearities allow DL models to capture mathematical relationships between the input features and labels. Over the years, DL has significantly impacted various science fields, such as computer vision, speech recognition, natural language processing, audio recognition, and bioinformatics [8]. DL allows computational models that use multiple processing layers composed of multiple non-linear transformations to learn data representations with numerous levels of abstraction. The application of DL has been highly relevant in medical imaging, particularly in image segmentation, denoising, detection, registration, and classification [9]. DL in the context of medical images makes use of pixel values rather than extracted or selected features. Hence, overcoming errors associated with inaccurate segmentation or subsequent feature extraction [5]. Automatically learning features from data and its self-learning capabilities are two of the most critical tools of DL.

The interest in DL stems from the evolution of convolutional neural networks (CNNs), a powerful way to learn valuable representations of images and other structured data. A CNN is a network architecture for DL which learns directly from data, eliminating the need for manual feature extraction. CNNs inspired by the biological structure of a visual cortex contain arrangements of simple and complex cells [10]. These cells are known to activate based on the subregions of a visual field. These subregions are called receptive fields. The neurons in a convolutional layer connect to the layer's subregions before that layer instead of being fully connected as in other types of neural networks. The neurons are unresponsive to the areas outside the image's subregions. They are the most popular neural networks for medical image analysis.

1.2 The Architecture of a Convolutional Neural Network

There are four CNN layers: the convolutional layer, the pooling layer, the rectified linear unit (ReLU) correction layer, and the fully connected layer.

1.2.1 The Convolutional Layer

The convolutional layer is the first and critical component of convolutional neural networks [11]. It serves as an extractor of features, as its purpose is to detect the representation of the data in the images received as input. A convolution converts all the pixels in its receptive field into a single pixel. The final output of a conventional layer is a vector.

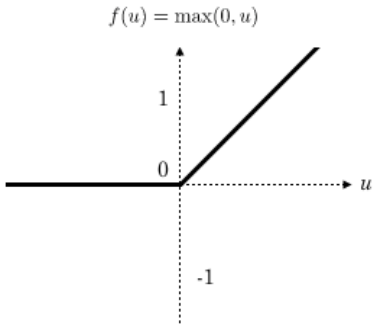
1.2.2 The Pooling Layer

The pooling layer often succeeds the convolutional layer. It receives several feature maps to create a new set of the same number of pooled feature maps. Pooling involves selecting a pooling operation, which reduces the images' size while preserving their essential characteristics. A window is selected to perform a pooling operation, then the input elements lying in that window are passed through a pooling function. The best advantage of the pooling layer is that it

reduces the number of parameters and introduces translation invariance [12]; if a small amount translates the inputs, the value for most pooled outputs does not change.

1.2.3 The ReLU Correction Layer

The ReLU correction layer refers to the real non-linear function defined by *ReLU(x) = max(0,x)* [11]. It is a piecewise linear function that will output the input directly if it's positive and zero otherwise.



It has become the most common activation function for many types of neural networks because of its advantages; Calculation of ReLU partial derivatives is much easier. It also does not allow gradients to disappear [12].

1.2.4 The Fully Connected Layer

The fully connected layer is usually the last. Here, all the inputs from each layer are connected to every activation unit of the next layer. It produces an output vector by applying a linear combination and possibly an activation function to the inputs received. The fully connected layer regulates the relationship between the location of the features in the image and a class. If the position of a feature at a certain point in the image is characteristic of a particular class, then the corresponding value in the table is given significant weight.

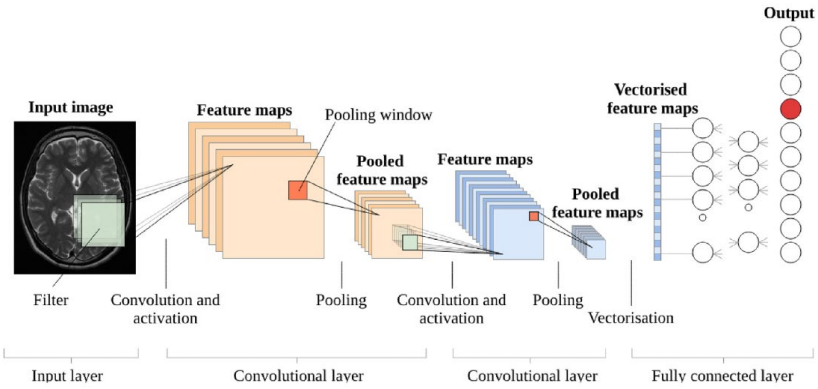


Figure 1.2: A representation of a CNN architecture showing convolutional layers, activation layer, pooling layer, and the fully connected layer [13].

1.3 Data Augmentation

DL for medical image analysis has often been constrained by the availability of labeled training data [14]. The performance (generalization) of a DL model is known to be associated with the volume of the training dataset. Hence, it is essential to help avoid overfitting our data and memorizing training sets by the DL models [15]. Data augmentation is a widely known technique that is cost-effective, less time-consuming and helps improve the generalization capabilities of deep neural networks. Thus, it can be perceived as an implicit regularizer. Data augmentation is used to increase the training dataset volume artificially. In medical image classification, the data augmentation technique is essentially done by employing transformations such as rotations, reflections, rescaling, translating, shearing, etc., to both the images and labels equally [14]. However, data augmentation can be achieved in two ways: offline augmentation and online augmentation.

- Offline augmentation: This form of augmentation consists of performing transformations on the images and saving the outputs on your pc storage. This increases the volume of the dataset by a factor equal to the number of transformations employed [16]

- Online augmentation: This form of augmentation consists of performing transformations on the mini-batches that would be fed to the model during training [16].

1.4 Transfer Learning

In medical imaging, there are cases where training data are difficult to collect or too expensive. A widespread assumption has been that training and test data must have identical feature spaces with the underlying distribution. However, this assumption does not usually hold for real-world data. Hence, models need to be built from the ground up if the features and distribution change [17]. Consequently, there is a need to create high-performance models trained that could transfer the knowledge learned from trained data across different domains [18]. This methodology is known as transfer learning. This technique reduces the dependence on many target domain data for constructing target learners. Transfer learning uses knowledge from one domain called the source to improve the learning performance or, sometimes, minimize the number of labeled examples required in a particular field of interest [19].

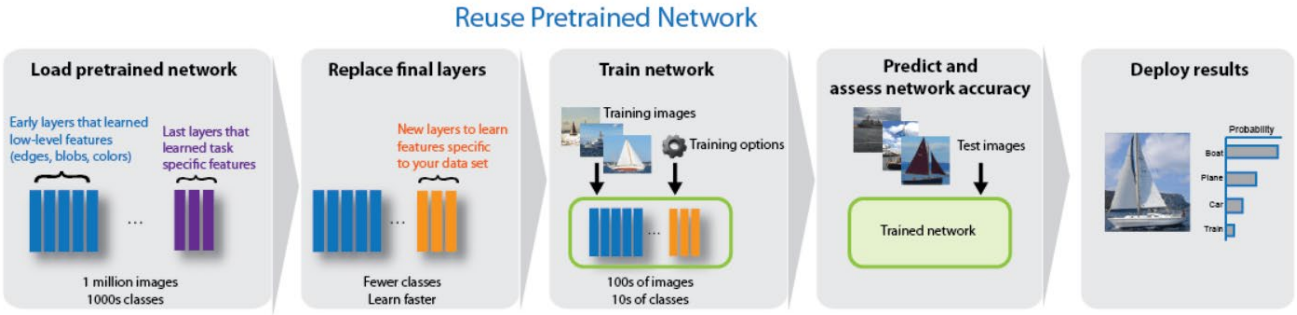


Figure 1.3: A schematic showing the use of a pre-trained network [20].

1.5 Feature Extraction

Dimensionality reduction is a significant aspect of data visualization, modeling, and analysis. It is a widespread preprocessing step in non-DL applications. One method of accomplishing this

task is feature extraction. Feature extraction refers to the process of converting raw data into numerical features that can be processed by a model while preserving the information contained in the original data. It leads to the concept of discovering outstanding features that are invariant to inconsequential transformations of the input [21]. Feature extraction is a more general method of transforming an input space onto a lower-dimensional subspace without losing the most relevant information [22].

Feature extraction is achieved either manually or automatically. Manual feature extraction requires identifying and describing the features that are deemed relevant for a particular task. This often calls for a good understanding of the dataset background. On the other hand, automated feature extraction uses deep networks or specialized algorithms to extract features from data without requiring manual intervention. With the rise of deep learning, feature extraction has been substituted mainly by the first layers of neural networks, especially for image data.

Many researchers have employed classical feature descriptors inspired by the success of natural image analysis which has yielded considerable success in histopathological image analysis [23]. The attributes of tumors in histopathological images are distinct from those in natural images. However, certain classical feature descriptors like local binary pattern (LBP), scale-invariant feature transformation (SIFT), and histogram of oriented gradient (HOG) [24] have had favorable outcomes in depicting histopathological images. L. Ladha et al. [25] outlined some advantages of feature extraction:

1. It helps in reducing the amount of redundant, irrelevant, or noisy data.
2. It increases the speed and accuracy of the model.
3. It improves the quality of the data.

4. It helps to understand or gain knowledge about the data.
5. It improves the generalization of the model.

1.6 Motivation For This Study

Brain tumors have been known to be difficult to treat, hence, having poor prognoses. Thus, our motivations for this study are listed as follows:

1. We would like to ascertain if there are potential insights that could be derived when mouse monoclonal antibodies targeting the isocitrate dehydrogenase (IDH1) is used to stain these tumor glioma cells.
2. When glioma cells grow from low grade to high grade, they become more diffusive and invasive. Hence, in this study, we would like to evaluate the performances of several DL and non-DL models employed in classifying the 5 tumor types of different grades.
3. We would also like to assess if our approach could potentially provide better therapy for patients diagnosed with these tumor types.

2 RELATED WORKS

2.1 Machine Learning Methods for Histopathological Images

With the evolution of ML in biomedical image analysis, numerous studies have been carried out on feature-based approaches for classifying histopathological images [6]. Kowal et al. in [26] focused on nuclei segmentation of 500 fine-needle biopsy images of breast cancer and extracted forty-two morphological, topological, and texture features. Afterward, these features were used to train different ML classifiers to classify them into benign and malignant classes. Osborne et al. [27] employed segmentation and morphology features, trained with a support vector machine (SVM) classifier for melanoma diagnosis in histopathological images. The suggested approach achieved an accuracy of 90%. Olgun et al. [28] proposed a method for classifying colon tissue histopathological images based on the local distribution of objects. This approach was then evaluated using an SVM that outperformed all thirteen classifiers with an accuracy of 93%. Muthu Rama et al. [29] utilized an SVM approach to classify oral mucosa histology images. Then a Bayesian classifier was implemented based on the defined space for characterizing inflammatory and fibroblast cells to observe the cells' distribution in a healthy state. M. Murat Dundar et al. [30] used expectation-maximization and watershed transformation to classify intraductal breast lesions. In this study, the system for automatic pre-invasive breast images was developed with 62 patient cases, and the overall classification accuracy was 87.9%. Filipczuk et al. [31] also carried out a nuclei segmentation of 737 cytology images of breast cancer, extracting twenty-five shape-based and texture-based features. Based on these features, four different ML classifiers – Naïve Bayes (NB), decision tree (DT), SVM, and k-nearest neighbor (KNN) – were trained for the classification of these cytological images into benign and malignant classes. Zhang et al. [32] combined LBP, the statistics from the gray-level co-

occurrence matrix, and curvelet transform to design a cascade random space ensemble system (with rejection options) for effective classification of microscopic biopsy images of breast cancer. Mazo et al. [33] proposed the classification of cardiac tissue into five classes using a patching approach that aims to optimize the path size to improve the representation. A cascade of linear SVMs separates the tissue into four distinct classes, after which a polynomial SVM classifies one of these four classes into two sub-classes.

Chan and Tuszynski [34] applied an SVM classifier to detect brain cancer using their fractal features. Their method achieved an F1 score of 97.9% when classifying histopathological images of magnification 40x into benign and malignant tumor classes. Furthermore, they achieved an F1 score of 56.5% by employing this method on a multi-class problem. Harai and Tanaka [35] suggested a colorectal computer-aided design (CAD) system to separate the nuclei, background, and stroma using an Otsu thresholding of the red channel. An SVM classifier was used to achieve an accuracy of 78.3% as opposed to a method based on texture features which achieved an accuracy of 67%. Zhang et al. [36] proposed a multi-scale classification that employed sparse encoding and Fisher's discriminant analysis to create a visual dictionary of SIFT features. This approach had an accuracy of 81.6% when an SVM classifier was used, performing more than the state-of-the-art method, which had an accuracy of 79.5%. Atupelage et al. [37] applied an SVM to fractal features. They aim to classify non-neoplastic and grade hepatocellular carcinoma histopathological images into five classes. This approach had a 95% accuracy outperforming the other methods that adopted texture features in the study. Vanderbeck et al. [38] applied an SVM to a combination of 413-dimensional feature vectors extracted from the white regions of liver histopathological images, classifying them into seven classes with an accuracy of 93.5%. Rahman et al. [39] analyzed texture abnormalities in oral squamous cell carcinoma using

histopathological samples. Histogram and Gray Level co-occurrence matrix (GLCM) extracted texture features from biopsy images from normal and malignant cells. Afterward, a linear SVM classifier was employed to classify oral cancer, which achieved 100% accuracy automatically. Spanhol et al. [40] published a breast cancer dataset called BreaKHis containing 7,909 histopathological images of 82 breast cancer patients. They used six different feature descriptors and 4 ML models – quadratic discriminant analysis (QDA), random forest (RF), KNN (with k=1), and an SVM with Gaussian kernel function. These 4 ML models were applied for binary classification of benign and malignant tumors. The accuracy was between 80 % and 85% using a 5-cross validation. Orlov et al. [41] carried out a comparison of three-color spaces (RGB, CIELAB, and grayscale) with Hematoxylin & Eosin (H&E) representation. They employed a weighted KNN, a radial basis function (RBF) network, and an NB classifier. The methods each had an accuracy of 99%, 99%, and 90%, respectively. Additionally, the best results were achieved for the color space termed eosin representation.

Several studies have shown that combining classifiers may enhance the performance of histopathological image classification. In their study, Kong et al. [42] classified neuroblastomas using textural and morphological features. They applied these methods using an ensemble approach combining KNN, linear discriminant analysis (LDA), NB, and an SVM classifier. They achieved an accuracy of 87.8% using the weighted voting rule. Zarella et al. [43] employed multiple classifiers in an ensemble of SVMs on ROIs segmented from a WSI. They were trained with subsets of features, achieving an accuracy of 88.6% when combined with a weighted sum (WS) function. Di Franco et al. [44] proposed an ensemble of SVM classifiers, trained with various images preprocessed by color spaces and Gaussian filters. Afterward, the classifiers were combined using the average rule, producing the best area under the curve (AUC) of about 97.8%.

Kruk et al. [45] used an ensemble of SVM and RF classifiers, trained with a subset of morphometric, textural, and statistical features extracted from nuclei. This method achieved an accuracy of 96.7%, outperforming a state-of-the-art and the best single SVM classifier, which both had an accuracy of 93.1% and 91.1%, respectively. Valkonen et al. [46], in their study, utilized Otsu, morphological operations, and histological constraints in segmenting a WSI. The ML models proposed for this study were RF, KNN, SVM, and logistic regression. The models were trained with textural, morphometric, and statistical features extracted from random patches of the segmented images. RF achieved the best accuracy of 93%. An ensemble of SVM and RF classifiers to classify prostate cancer was employed by Gertych et al. [47]. The SVM was used to separate the stroma and epithelium, while the RF classifier was used to identify the benign and malignant tissues. The best accuracy was 68.4% for cancer detection. Romo-Bucheli et al. [48] proposed an ensemble of an Adaboost classifier for grading skin cancer. This method was used to classify images described by features representing nuclei distribution created using graph theory. The ensemble achieved an accuracy of 72%.

2.2 Deep Learning Methods for Histopathological Images

The adoption of DL techniques in medical imaging has increased due to the positive impact observed on various tasks. Many recent studies have employed DL methods to classify histopathological images with and without leveraging transfer learning. DL has achieved tremendous performance in many digital pathology tasks. Malon et al. [49] were among the first set of authors that applied DL to histopathological images. They employed a classical LeNet-5, a 7-layered CNN architecture initially proposed by Lecun et al. [50] in 1998, to learn a representation of histopathological images segmented by a support vector regression (SVR) model. An SVM was applied to the features extracted by the CNN in [49] to find the mitotic

nuclei. Dev Kumar et al. [51] proposed a method for automatically identifying relevant ROIs from oral tissue histological images to detect oral squamous cell carcinoma. They employed a 12-layered deep CNN to different segmented keratin, epithelial, and subepithelial regions. Keratin pearls were detected from the segmented keratin region with the texture-based feature (Gabon filter). Then an RF was applied to classify the features attaining an accuracy of 96.88%. Sharma et al. [52] used an AlexNet CNN, as well as other custom CNN architectures, to classify benign and malignant tumors. Because of the small sample size, they also had to perform data augmentation by applying patching and affine transformations. Both AlexNet and other custom CNNs performed favorably to most handcrafted features. In their study, Khosravi et al. [53] evaluated the performance of different CNN models, including Inception and ResNet, as well as the combination of both on eight different datasets. Some of the datasets include breast, lung, and bladder tissues stained with H&E. The results, despite having been applied to raw images without preprocessing, had satisfactory performances. Zerhouni et al. [54] suggested a wide residual CNN in the classification of mitotic and non-mitotic pixels in breast histopathological images from the MICCAI TUPAC challenge dataset. The CNN was then trained on mitotic and non-mitotic patches extracted from the ground truth images. The results showed that the method employed in the study outperformed most other approaches.

Kainz et al. [55] presented two CNNs based on the LeNet-5 model architecture for the segmentation and binary classification of benign and malignant colorectal cancer gland tissue. The first CNN model separates glands from the background while the other identifies gland-separating structures. The experimental results on Warwick-QU colon adenocarcinoma and GlaS@MICCAI2015 challenge datasets showed a gland tissue classification accuracy for both models as 98% and 95%, respectively. AlexNet was adopted by Stanitsas et al. [56] to classify

breast cancer histopathological images. The performance of this CNN was compared against some handcrafted feature extractors and shallow classifiers. The result showed that the CNN model did not perform better than the shallow methods. Spanhol et al. [57] evaluated architectures based on AlexNet to classify the dataset in [40]. The result showed that depending on the magnification at the pixel level, the CNN model achieved average accuracy rates between 81.7% and 88.6%. The DL method outperformed other non-DL approaches applied in the study. Li et al. [58] evaluated the performance of an SVM model on handcrafted features and features extracted from DL models (AlexNet and Inception-V1) to classify regions of colon histology images as gland or non-gland. The combination of handcrafted features with SVM achieved a better result. Talo [59] employed ResNet-50 and DenseNet-161 to classify grayscale and color histopathological images. With an accuracy of 95.79% for ResNet-50 and 97.77% for DenseNet-161, these results outperformed the existing studies in the literature. Kwak and Hewitt [60] compared the performance of a proposed 6-layer CNN to other CNNs (AlexNet, VGG, GoogLeNet, and ResNet) and ML classifiers (SVM, KMM, RF, and NB) in identifying prostate cancer. The proposed CNN had an AUC of 0.974, which outperformed all other methods applied in the study. Budak et al. [61] employed an end-to-end model based on ALexNet and a bidirectional long-short term memory (BLSTM) identifying breast cancer cells in the BreaKHis dataset. The convolutional layers were used to encode the images, flattened before being fed to the BLSTM. The result showed that the proposed model achieved an accuracy of 95.69%, 93.61%, 96.32%, and 94.29% for the magnification factor of 40x, 100x, 200x, and 400x respectively.

Han et al. [62] suggested a class structure based on deep CNN for breast multi-classification using hierarchical feature representation. They achieved an accuracy of 95.9%. Nawaz et al. [63]

employed a DenseNet model for the multi-classification of breast cancer in the BreaKHis dataset to predict the subclass of breast cancer tumors. They achieved an accuracy of 95.4%. Ciresan et al. [64] presented a method that won the 2012 ICPR Mitosis Detection Contest. They employed a deep CNN in detecting mitosis in breast histology images. The CNN was utilized as a pixel-wise classifier through a sliding window manner in detecting mitosis. In other words, it was trained to classify pixels in the image from a patch centered on the pixel. This study would lay the foundation for other studies. For example, Sirinukunwattana et al. [65] used a spatially constrained CNN (SC-CNN) to classify and detect nuclei in histopathological images. More accurately, they employed SC-CNN in estimating the likelihood of a pixel being the center of a nucleus. They also utilized a neighboring ensemble predictor with CNN to predict the label of a detected cell nucleus. Xu et al. [66] employed a stacked autoencoder (SAE) for detecting breast cancer in histopathological images. They adopted a denoising autoencoder to improve its robustness to outliers and noise in training the DL model. Yang et al. [67] developed a 7-convolutional layers CNN to classify kidney cancer in 500 histopathological images as tumor or non-tumor. They achieved an accuracy of 98%. Wang et al. [68] adopted the combination of CNN and handcrafted features for detecting mitosis in breast histopathological images. They presented a cascaded approach that probably maximizes exploiting two distinct feature sets. This approach demanded less computational cost and achieved greater accuracy. They also utilized crowdsourcing in the learning process of the CNN to exploit additional data sources annotated by non-expert users for mitosis detection [69]. They used different image scales to train a multi-scale CNN to perform mitosis detection and provide the crowds with mitosis candidates for annotations. These annotations were then fed to the existing CNN for model refinement and ground-truth generation. Guo et al. [70] proposed a two-stage approach in their study. They

employed an Inception-V3 for classifying the tumor regions of breast cancer WSI pathological images, followed by a cascaded deep CNN for refined segmentation.

Yan et al. [71] proposed a novel hybrid convolutional and recurrent deep neural network for breast cancer histopathological image classification in their study. They used this proposed method to classify the images into normal, benign, in situ, or invasive carcinoma. This approach consists of dividing the histopathological images into 12 patches. Then, using fine-tuned Inception-V3 to extract the features from the patches. Based on the richer multilevel feature representation of the patches, this method combines the advantages of the convolutional and recurrent neural networks while preserving the spatial correlations between the patches. The experimental results showed that this method achieved an accuracy of 91.3% for the 4-class classification task, outperforming the state-of-the-art approach. De Matos et al. [72] suggested a classification approach that uses Inception-V3 to extract features from breast cancer histopathological images. The method improved the accuracy by 3.7% and an additional 0.7% using the irrelevant patch elimination. Albarquoni et al. [69] introduced a framework that learns from crowds that handle data aggregation as part of the learning process of a CNN via an additional crowdsourcing layer for challenging classification tasks. This added layer was augmented to the CNN to aggregate ground truth from the crowd vote matrix. Experimental results on the AMIDA13 dataset indicated that the proposed model architecture was robust to noisy labels and positively improved the performance. Vizcarra et al. [73] developed a classification image pipeline using the BACH dataset that combines a shallow learner (SVM) and CNN for breast cancer histology classification. The pipeline consists of extracting speeded-up robust features (SURF) for the SVM, color normalization (Reinhard method), and image resizing for both Inception-V3 and Inception-ResNet-V2. The results showed an average

accuracy of 79% and 81% for the shallow learner and CNN, respectively. When fused with both algorithms, the system obtained an accuracy of 92%, outperforming any individual learner. Brancati et al. [74] adopted a DL approach for two different cases: the detection of invasive ductal carcinoma in breast histological images and the classification of lymphoma subtypes. For the first case, the convolutional layers were trained without supervision to learn a latent representation to reconstruct the input image. For the other case, the fully connected layers were trained while supervised. However, both cases addressed the challenges by adopting a residual CNN, which is part of a convolutional autoencoder network. The performances were evaluated on the public dataset of digital histological images and compared with those obtained using different deep neural networks (UNet and ResNet). The experimental results showed an improvement of 5.06% in the F-measure score for the detection task and 1.09% in the accuracy measure for the classification task.

Ataky et al. [75] proposed a novel approach for augmenting histopathological image datasets and distributing inter-patient variability through image blending using the Gaussian-Laplacian pyramid. Both sides of the histopathological images are joined on each level of the Laplacian pyramid. Also, from the joint pyramids, the original image is reconstructed. This constitution combines the stain variation of two patients, avoiding the color differences that may mislead the learning process. Experimental results with a texture CNN [76] have shown promising gains vis-à-vis most data augmentation techniques on the BreaKHis dataset presented in the literature. Ciompi et al. [77] proposed a colorectal cancer tissue classification system based on an 11-layer CNN. In this study, the authors investigated the importance of stain normalization in tissue classification of colorectal cancer tissue samples in H&E-stained images. Experimental results on the colorectal cancer dataset validated the proposed CNN's performance and the role of stain

normalization in colorectal cancer tissue classification. Sheikh et al. [78] proposed a four-input 24-layer custom CNN to classify histopathological images. This proposed model fuses multi-resolution hierarchical feature maps at different layers, learning different scale image patches to account for cells' overall structures and texture features. Experimental results on ICIAR2018 and BreaKHis datasets showed that the proposed model outperformed existing state-of-the-art. The explanation of our dataset and proposed methodologies will be discussed comprehensively in Section 3.

3 MATERIAL AND METHODS

Brain tumors are difficult to treat and have poor prognoses. One main reason is that they are invasive without a clear boundary between tumor and normal brain tissue. Consequently, it is very challenging to detect and classify. Glioma is a tumor that starts at the brain or spine. It makes up about 80% of all malignant brain tumors.

This section introduced our dataset, its preprocessing methodology, training, validation, testing criteria, and the data augmentation process. We also discussed the feature extraction techniques we employed and, finally, the Bayesian optimization of our non-DL models.

3.1 Data Collection

We collected 125 glioma slides, each corresponding to one patient. Sections of human gliomas of grades II-IV were stained with mouse monoclonal antibodies targeting the IDH1 RI32H mutation. Afterward, they were scanned using Olympus Nanozoomer whole-slide scanner at a resolution of 40x. Amongst the slides includes: 29 astrocytoma grade II (AII), 20 astrocytoma grade III (AIII), 9 oligoastrocytoma grade II (OAII), 7 oligoastrocytoma grade III (OAIII), 29 oligodendroglioma grade II (OII), 26 oligodendroglioma grade III (OIII), 4 glioblastoma multiforme (GBM), and 1 secondary glioblastoma multiforme (GBMII). About half of the images are low grad gliomas, including AII, OAII, and OII.

3.2 Preprocessing

Mutation in IDH1 occurs in up to 75% of gliomas. The dataset used in this study contains histopathological images of brain tumors stained with mutation-specific antibodies for IDH1 [79]. All works were carried out on MATLAB version 2022a. Each folder containing the brain tumor classes was saved in a single folder. The ImageDatastore method was used to store the

collection of each folder class and then imported to MATLAB However, we did not initiate any stain normalization process in this study and thus used the original images.

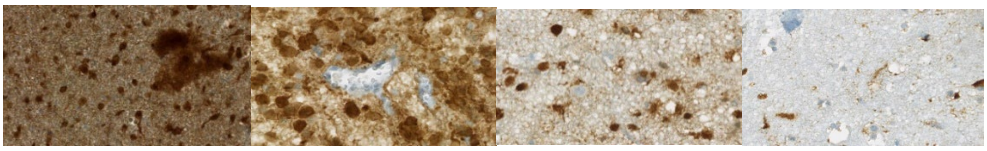


Figure 3.1: Color variations of images in our data.

The countEachLabel method was then used to count and keep track of each label in the datastore. Afterward, the data were split into training, validation, and test using the splitEachLabel method on MATLAB. The learnable layers and the classification layer of the transfer learning DL model were adjusted to reflect the weights and number of classes of our data. The imageDataAugmenter was then used to specify the data augmentation parameters before using the augmentedImageDatastore method to transform the batches of training and validation data with optional preprocessing such as resizing to fit the network and the specified parameters from the imageDataAugmenter. During training, the epoch was set to 10, a stochastic gradient descent optimizer (sgdm) was selected for optimization, and the learning rate was set as 0.0001.

3.3 Training Criteria

For the individual DL models, we selected 60% of data for training, 20% for validation, and 20% for testing. This way, we had 1440 images for training, 480 for validation, and 480 for testing. On the other hand, for the individual non-DL models, we selected 70% of data for training and 30% for testing. In this way, we had 1680 images for training and 720 images for testing. The statistics about training, validation, and testing for both types of models are shown below.

Table 3.1: Criteria for selection of training, validation, and test for DL classification

	No. of Images	**Percentage**
Training	1440	60
Validation	480	20
Test	480	20
Total	2400	100

Table 3.2: Criteria for selection of training and test for our non-DL classification

	No. of Images	**Percentage**
Training	1440	60
Test	480	40
Total	2400	100

3.4 Data Augmentation

In this study, we employ the online augmentation technique. By applying the imageDataAugmenter method in MATLAB library to our data, we generate batches of image data with real-time data augmentation. This way, we ensure that our network sees data variations at each epoch during the training process. Firstly, input as a batch of images is fed to the imageDataAugmneter, which then transforms each image in the batch by a series of random translations, rotations, reflections, etc. The rotation we specified has a range of 45, corresponding to a random rotation angle between [-45, 45] degrees. We also set the translation range as 3, corresponding to a random translation between [-3, 3]. Then we set the vertical and horizontal reflection to be true. Each image is reflected vertically and horizontally with a 50% probability.

Finally, the randomly transformed batch is returned to the calling function. All these parameters, alongside their values, are shown below.

Table 3.3: Parameters of data augmentation

Parameters of Image augmentation	**Values**
RandXReflection	True
RandYReflection	True
RandRotation	45
RandXShear	30
RandYShear	30
RandXTranslation	3
RandYTranslation	3

3.5 Feature Extraction Techniques

In this study, we employed two techniques for feature extraction: HOG features extraction and feature extraction with CNN.

3.5.1 HOG Features Extraction

A histogram of gradient (HOG) is a feature descriptor for object detection. It focuses on the structure or the shape of an object. For the regions of the image, it generates histograms using the magnitude orientations of the gradient. Suppose *f(x, y)* records the color of the pixel at location (x, y), and the gradient vector of the pixel (x, y) according to [80] is defined as

$$\nabla f(x,y) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{df}{dx} \\ \frac{df}{dy} \end{bmatrix} = \begin{bmatrix} f(x+1,y) - f(x-1,y) \\ f(x,y+1) - f(x,y-1) \end{bmatrix}$$

The $\frac{df}{dx}$ term is the partial derivative on the x-direction, computed as the color difference between the adjacent pixels left and right of the target, *f(x + 1, y) – f(x-1, y).*

Similarly, the $\frac{df}{dy}$ term is the partial derivative on the y-direction, computed as the color difference between the adjacent pixels above and below the target, *f(x, y+1) – f(x, y-1).*

Then, the magnitude and the angle as stated in [80] are

$$\text{Magnitude, } \mu = \sqrt{g_x^2 + g_y^2}$$

$$\theta = \arctan\left(\frac{g_y}{g_x}\right)$$

After calculating the magnitude, the gradient matrices (magnitude and angle matrix) are divided into CxC cells to form a block. For each block, a 9-point histogram is calculated. A 9-point histogram develops a histogram with 9 bins [81]. With few bins, a pixel whose orientation is close to a bin boundary might end up contributing to a different bin, were the image to change slightly. To prevent these quantization artifacts, each pixel in a cell contributes to two adjacent bins (modulo B), a fraction of the pixel's magnitude that decreases linearly with the distance of that pixel's gradient orientation from the two bin centers.

According to [82], the bins are numbered 0 through B – 1 and have width $w = \frac{180}{B}$. Bin *i* have boundaries $[w_i, w_{(i+1)})$ and center $c_i = w_{(i+\frac{1}{2})}$. A pixel with magnitude μ and orientation θ contributes a vote

$$\upsilon_j = \mu\frac{cj+1-\theta}{w} \text{ to bin number } j = \left[\frac{\theta}{w} - \frac{1}{2}\right] \mod B \qquad \{\text{for the jth bin}\}$$

As sated in [81], similarly,

$$\upsilon_{j+1} = \mu\frac{\theta - cj}{w} \qquad \{\text{for the (j+1)th bin}\}$$

Once the histogram computation is done for all blocks, the cells are grouped into overlapping blocks of 2 x 2 cells each [81] so that each block has a size 2C x 2C pixels. Two horizontally or vertically consecutive blocks overlap by two cells, that is, the block stride is C pixels. Consequently, each internal cell is covered by four blocks. Concatenate the four-cell histograms in each block, as shown in [82], into a single block feature **b** and normalize the block feature by its Euclidean norm:

$$\mathbf{b} \leftarrow \frac{b}{\sqrt{||b||^2 + \epsilon}}$$

In this expression ϵ is a positive constant that does not allow division by zero in gradient-less blocks. The evidence for choosing this scheme over others is entirely empirical.

Block normalization is a compromise: On one hand, cell histograms need to be normalized to reduce the effect of changes in the contrast between images of the same object. On the other hand, overall gradient magnitude does carry some information, and normalization over a block – a region frater than a single cell – preserves some of this information, namely, the relative magnitudes of gradients in cells within the same block. For each cell covered up to four blocks, each histogram is represented up to four times with up to four different normalizations.

The normalized block features are connected into a single HOG feature vector **h**, which is normalized as follows:

$$\mathbf{h} \leftarrow \frac{h}{\sqrt{||h||^2+\epsilon}}$$

$$h_n \leftarrow \min(h_n, \tau)$$

Here, h_n is the n-th entry of **h** and τ is a positive threshold ($\tau = 0.2$) [82]. Clipping the entries of h so that it can be no larger than τ (after the first normalization) ensures that huge gradients do not have too much influence – they would end up washing out all other image detail. The final normalization makes the HOG feature independent of overall image contrast.

Visualizing HOG Features

The figure below depicts the process of HOG feature extraction on an image for different cell sizes.

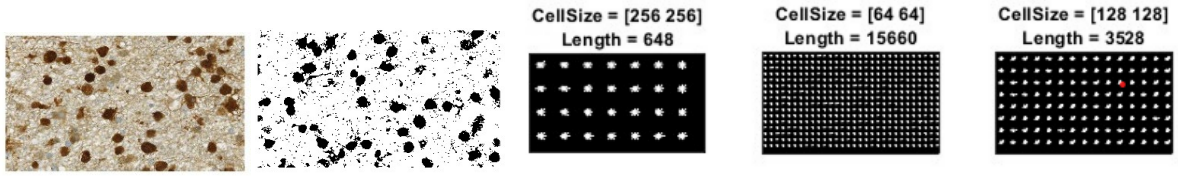


Figure 3.2: A representation of different cell sizes of HOG features extracted from an image

3.5.2 Feature Extraction With CNN

In this study, we also employed using ResNet-18 in extracting the features from our image datasets for machine learning applications. After splitting our image datasets into training and test, we resize them to the size suitable for our CNN model (224x224x3) using the augmentedImageDatastore in the MATLAB library. Each layer of the CNN produces a response

to an input. However, just a few layers within the CNN are suitable for image feature extraction. As visualized below, we use the initial layers to capture features such as edges and blobs.

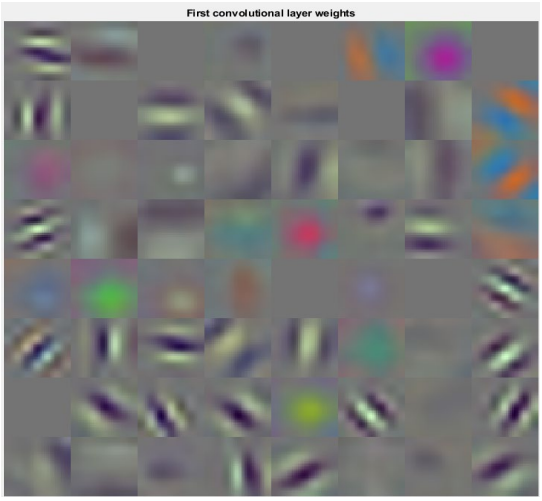


Figure 3.3: First convolutional weights

The figure above indicates that the model has learned filters for capturing these features. The primitive features are then combined to form higher-level image features. Afterward, we selected the layer (pool5) before the fully connected layer to extract our features using the activations method. The activation output is specified as rows to make it suitable for our non-DL classification models.

3.6 Bayesian Optimization

In this study, we also applied bayesian optimization to our non-DL models. Bayesian optimization is an approach that uses the Bayes theorem to minimize a scalar objective function *f(x)* for x in a bounded domain [83]. The Bayes theorem is an approach for calculating the conditional probability of an event, as depicted in [84]

$$P(A \mid B) = P(B \mid A) * P(A) / P(B)$$

By normalizing value of *P(B)*, this equation becomes

$$P(A \mid B) = P(B \mid A) * P(A)$$

This conditional probability referred to as the *posterior probability* given as

$$\text{Poserior} = \text{likelihood} * \text{prior}$$

This theorem provides the schema that can be used to quantify the assumptions about an unknown objective function given samples from the domain and their evaluation [84]. In other words, the posterior probability is a surrogate objective function. A surrogate objective function is the approximation of an objective function.

Thus, Bayesian optimization builds a probability model of the objective function and uses it to select hyperparameters to evaluate the true objective function [85].

4 RESULTS

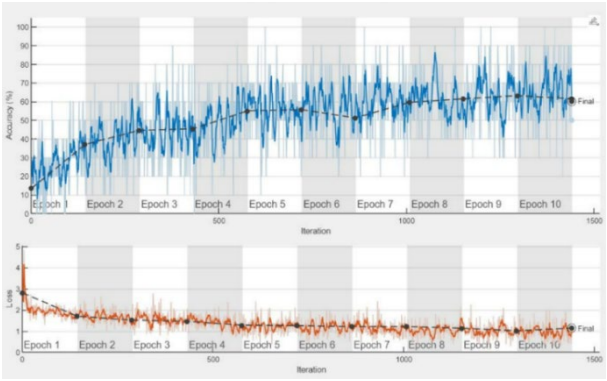
In this section, we describe and compare the results of eight different class of brain tumor histopathological images classification/recognition using 5 DL approaches. Afterward, we reiterate these process combined with an ensemble classifier and 5 non-DL approaches on a reduced dataset consisting of five classes to evaluate the performances of each model.

Table 4.1: The Transfer Learning DL classifiers used for the brain tumor classification

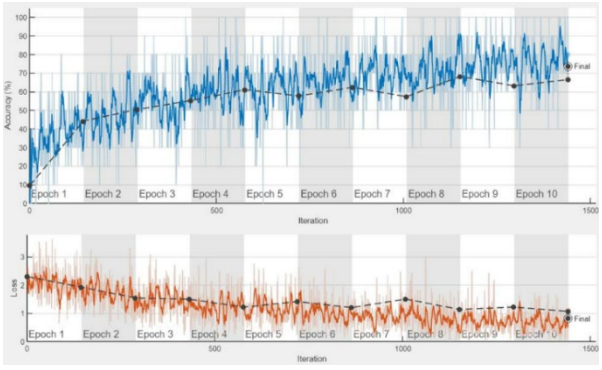
Transfer Learning DL Model	**Network Deep (Layer Size)**	**Parameters (Millions)**	**Image Input Size**
GoogLeNet	22	7.0	224 x 224
ResNet-50	50	25.6	224 x 224
Inception-V3	48	24.0	299 x 299
DenseNet-201	201	20.0	224 x 224
Xception	71	23.0	299 x 299

4.1 Comparison of Brain Tumor Histopathological Images Using the Complete Dataset

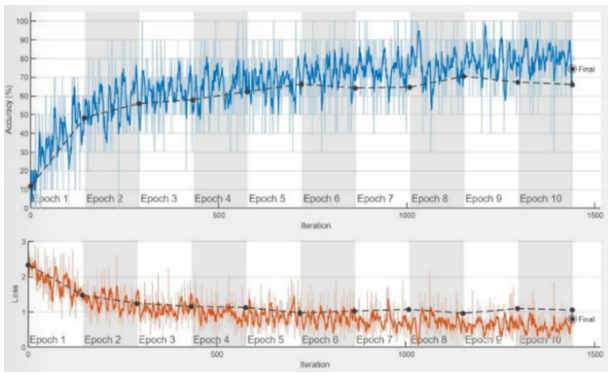
Here, we analyze the performance of 5 DL models using the complete dataset of tumor images.



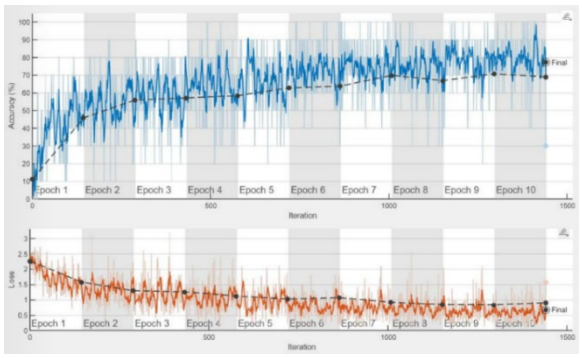
(a) GoogLeNet



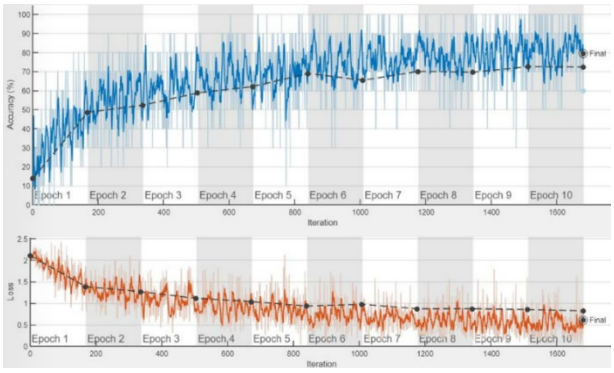
(b) ResNet-50



(c) Inception-V3



(d) DenseNet-201



(e) Xception

Figure 4.1: The validation and training plots for the 5 DL models and their loss plots for the complete dataset

Fig. 4.1 shows that the validation and training loss for all models, excluding (a) failed to converge. This might indicate that these models may not generalize well with new sets of data due to signs of overfitting. The erractic patterns reflects the low volume of dataset currently available for this study

Table 4.2: Accuracy for each Transfer Learning DL model for the complete dataset

Transfer Learning DL	**Accuracy (%)**
GoogLeNet	68
ResNet-50	77
Inception-V3	79
DenseNet-201	80
Xception	79

Table 4.2 shows that the best performing model was DenseNet-201 which achieved the best classification accuracy at 80% followed by Inception-V3 and Xception both tied at 79%. The worst performing model GoogleNet at 68% despite showing best indication of a genearlized model for the dataset.

Although the table above shows the average performance of each DL algorithm, it does not provide information about the weakness of each model and what accounts for misclassifications in this study. We generate the confusion matrices shown in Fig. 4.2 to evaluate our model performance for each class of brain tumor.

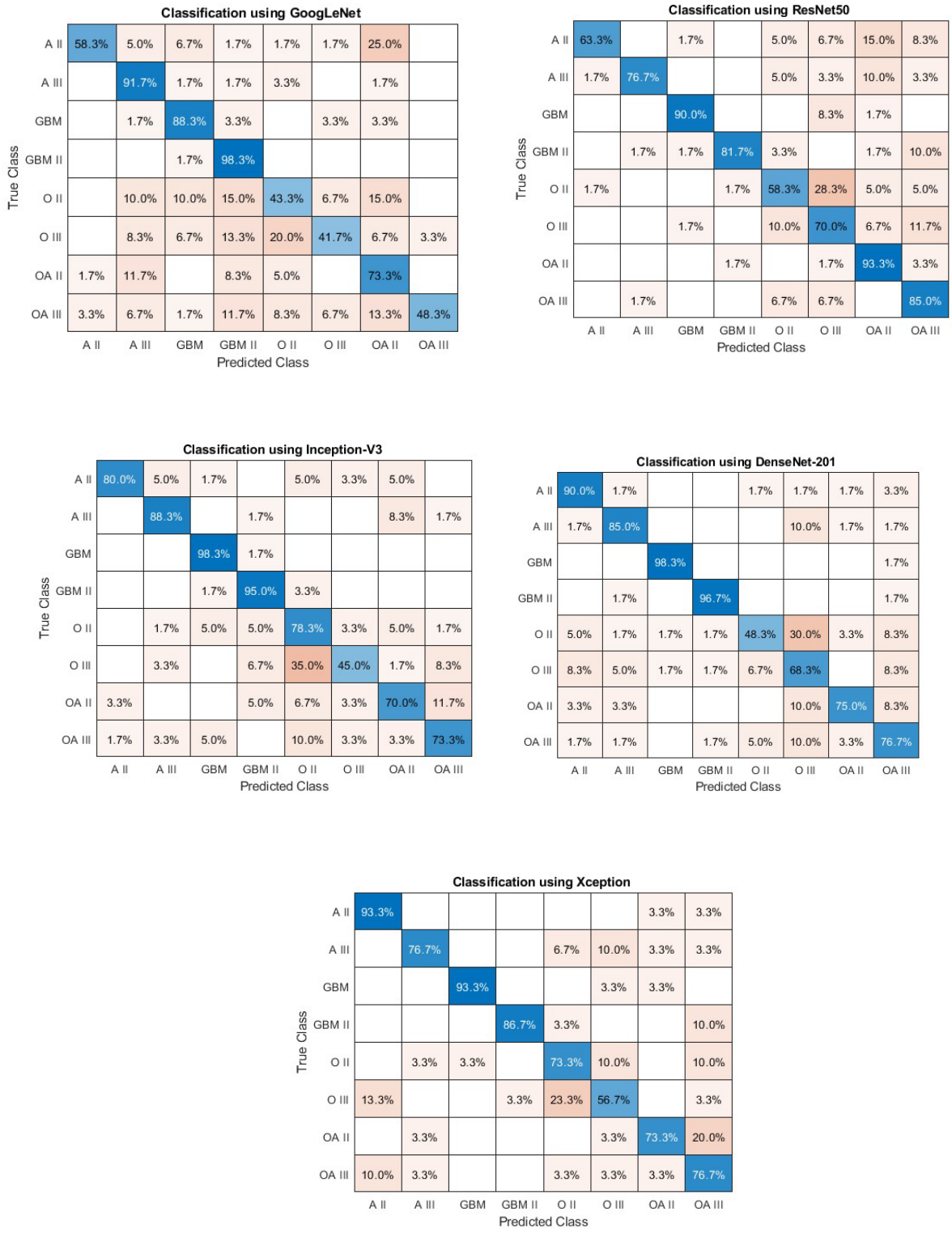


Figure 4.3: The confusion matrices of the best and worst DL model for the complete dataset

We designed a summary table in Table to briefly describe the classes classified with the highest rate and the lowest rate for each model. The table also described the classes that had the highest misclassified rates for each model

Table 4.3: *Summary of the DL confusion natrices for the complete dataset*

DL Models	**Best Classified class**	**Worst Classified Class**	**Most Misclassified Class**
GoogLeNet	GBMII	OII	OAII
ResNet-50	OAII	OII	OIII
Inception-V3	GBM	OIII	OII
DenseNet-201	GBM	OII	OIII
Xception	GBM and AII	OIII	OII

GBM had the highest classification rate with an average rate of 93.6% while GBMII was the second-highest with an average classification rate of 91.7%. OII had the highest misclassification rate followed by OIII. OII was misclassified as OIII with an average error rate of 19%. On the other hand, OIII was misclassified as OII with an average error rate of 15.7%.

4.2 Formation of the Reduced Dataset

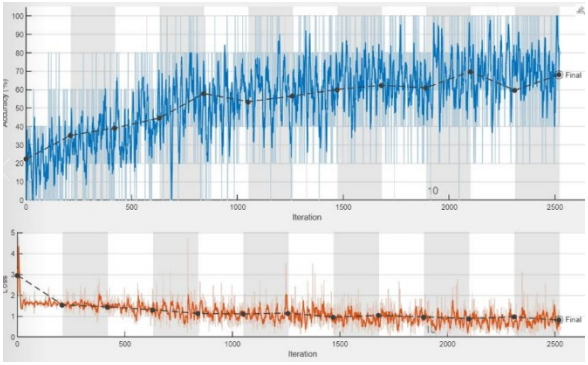
To further test if we can improve the accuracy of each model, we discarded the OA class known to have features of both oligodendroglioma and astrocytoma. We also combined the GBM and GBMII classes to form one glioblastoma multiforme (GBM) class. Hence, we now have 5 classes, which are

- Astrocytoma (grade II)
- Astrocytoma (grade III)

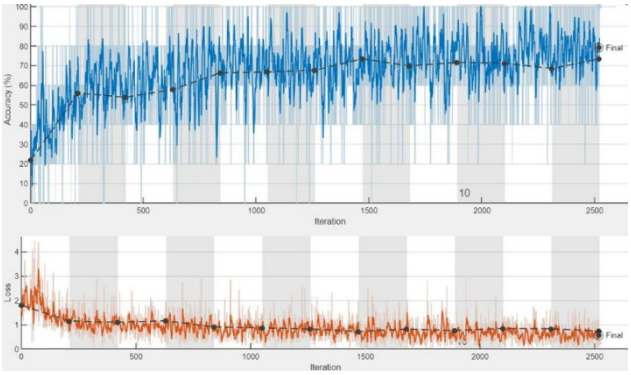
- Oligodendroglioma (grade II)
- Oligodendroglioma (grade III)
- Glioblastoma

4.2.1 Comparison of Brain Tumor Histopathological Image Classification Using the Reduced Dataset

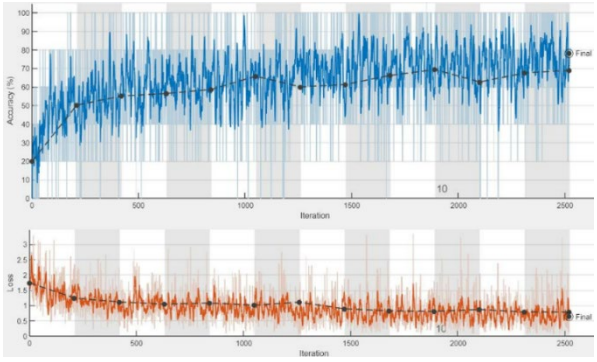
Similarly, we analyzed the performance of the 5 DL models using the reduced dataset of brain tumor images.



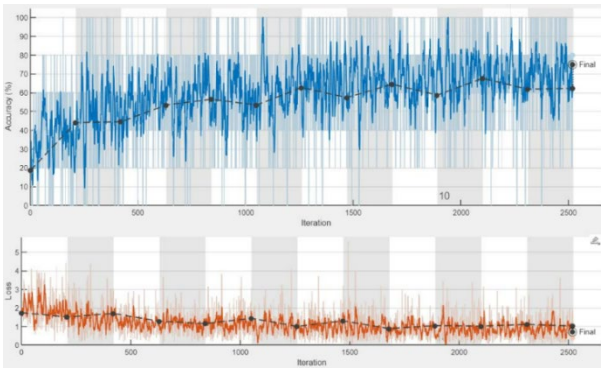
(a) GoogLeNet



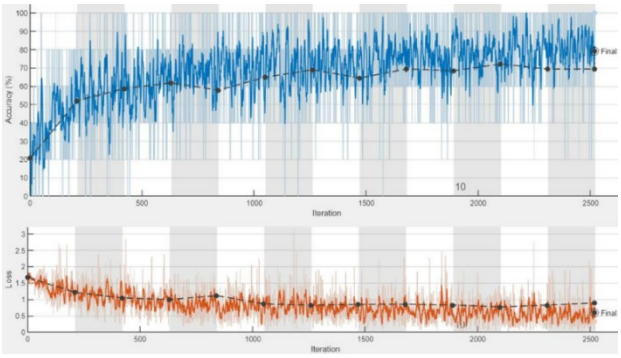
(b) ResNet-50



(c)Inception-V3



(d) DenseNet-201



(e) Xception

Figure 4.4: *The validation and training plots for the 5 DL models and their loss plots for the reduced dataset*

Fig. 4.3 shows that the validation and training loss for all models converged. This might indicate that these models may generalize well with new sets of data.

Table 4.4: *Accuracy for each Transfer Learning DL model for the reduced dataset*

Transfer Learning DL	**Accuracy (%)**
GoogLeNet	76
ResNet-50	79
Inception-V3	80
DenseNet-201	82
Xception	76

Table 4.4 shows that the best performing model was also DenseNet-201 which achieved the best classification accuracy at 82% followed by Inception-V3 at 80%. There was no significant difference in terms of accuracy for all models except GoogLeNet. Xception achieved a lower accuracy compared to the its accuracy for the complete daatset.

Similarly, confusion matrices were generated in Fig. 4.4 to assess each DL model performance for each of the bain tumor class

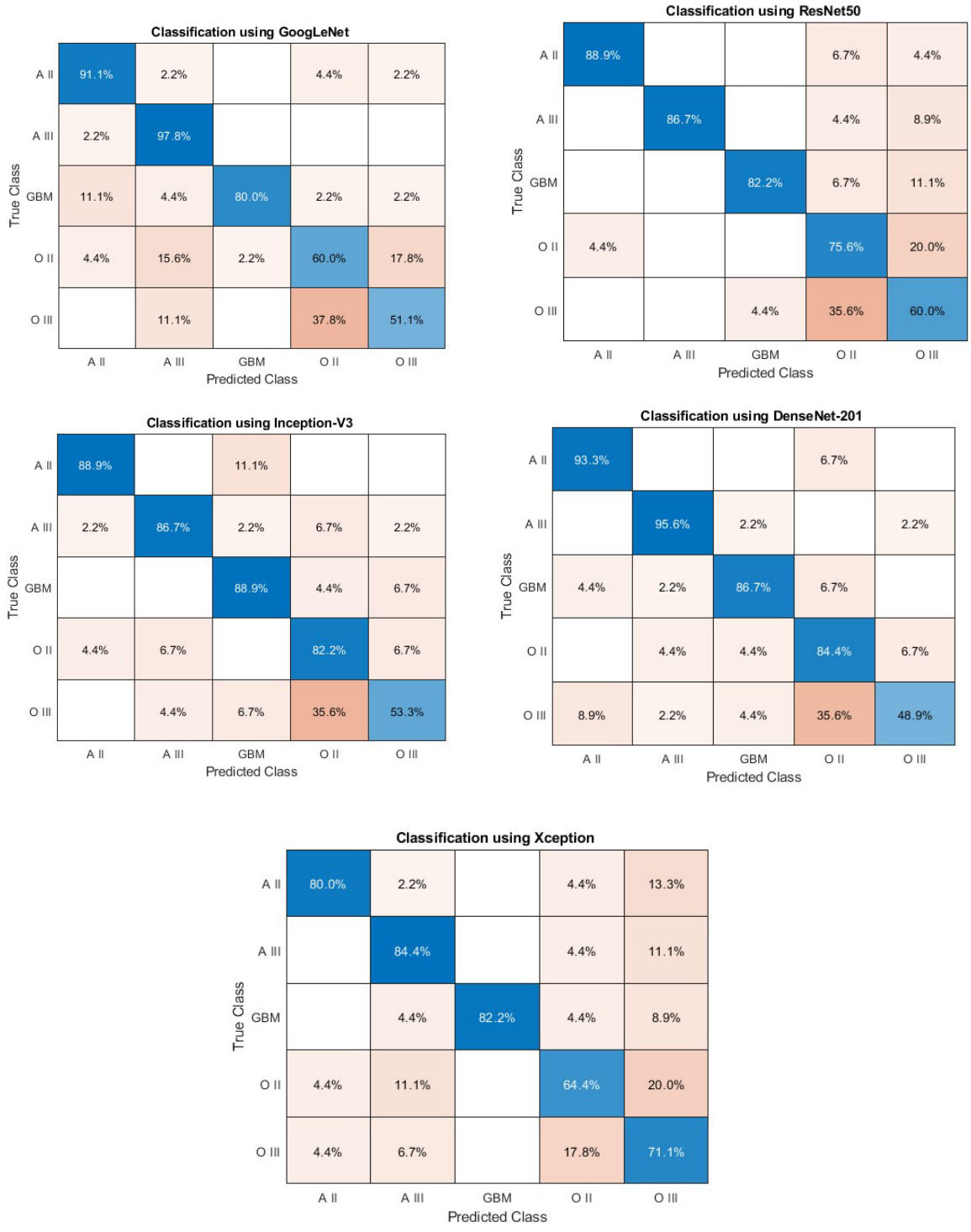


Figure 4.5: The confusion matrices of the best and worst DL model for the reduced dataset

Table 4.5: Summary of the DL confusion matrices for the reduced dataset

DL Models	Best Classified Class	Worst Classified Class	Most Misclassified Class
GoogLeNet	AIII	OIII	OII
ResNet-50	AII	OIII	OII
Inception-V3	AII and GBM	OIII	OII
DenseNet-201	AIII	OIII	OII
Xception	AIII	OII	OIII

AIII had the highest classification rate with an average rate of 90.2%, followed by AIII with an average rate of 88.4% and then GBM. OII had the highest misclassification rate. It was often misclassified as OIII with an average error rate of 32.5%.

4.2.2. HOG Features with Ensemble Method

A cell size of 256 x 256 was chosen as the best compromise. This cell size encodes enough spatial information to visually identify an image shape while limiting the numbe of dimensions in the HOG feature vector. An ensemble method (Bag) was selected as the best classifier after initializing an automated machine learning process with bayesian optimization.

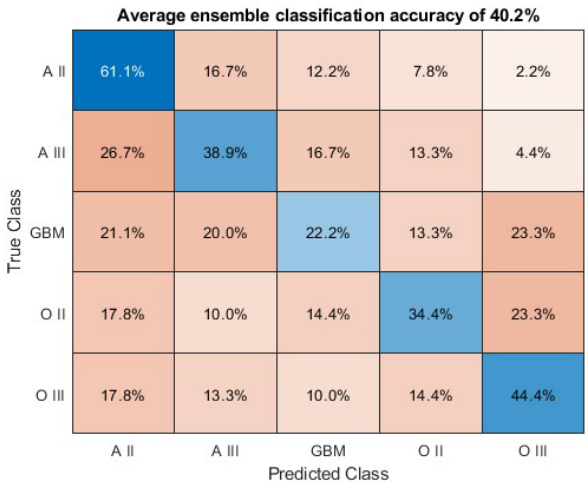
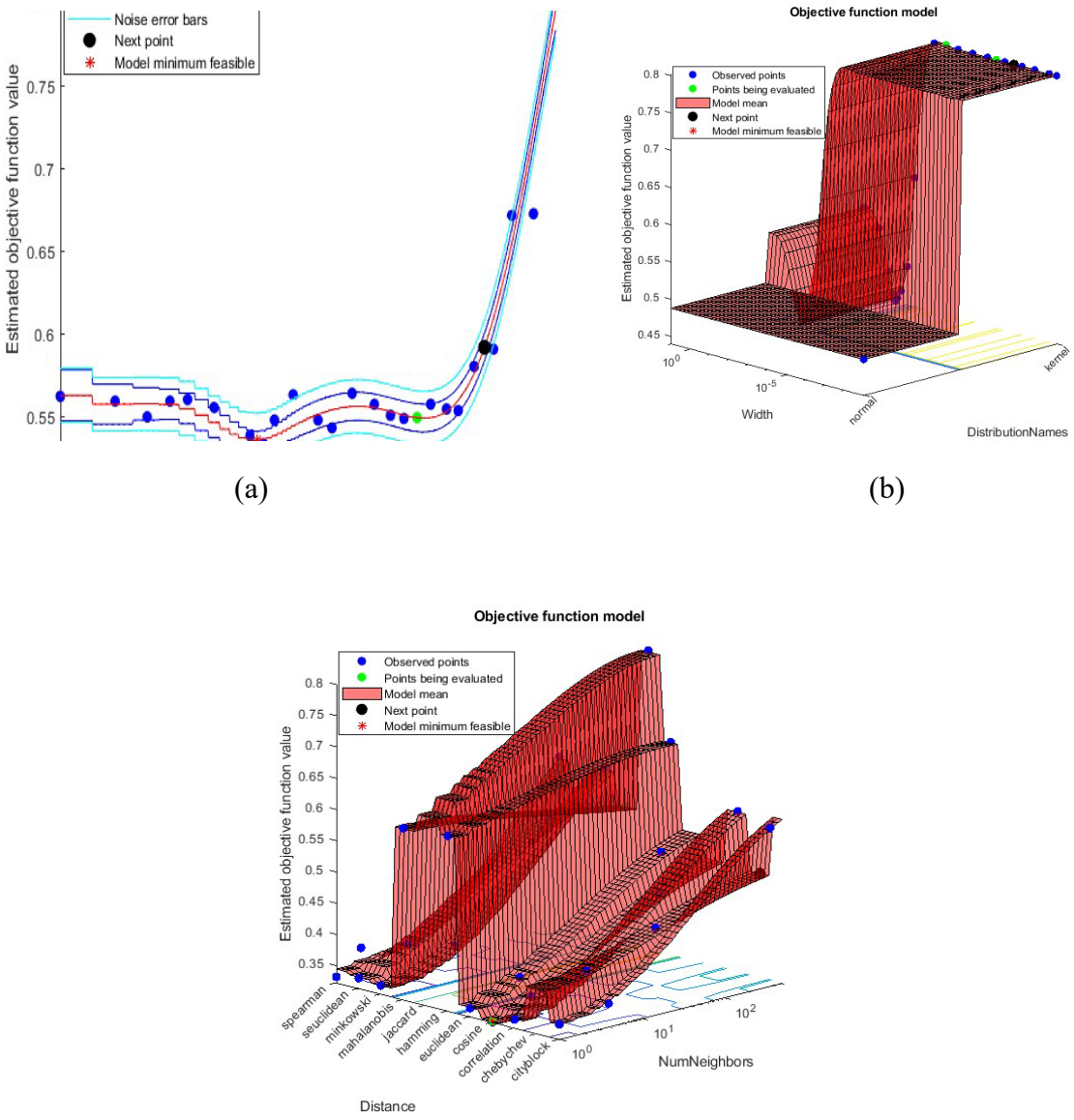


Figure 4.6: *confusion matrix for the bagging of decision trees (ensemble) classifier.*

This method achieved a classification accuracy of 40.2%. AII had the highest classification rate at 61.1%. GBM had the lowest classification rate at 22.2%. AII was misclassified as AIII with an error rate of 26.7% while OIII was often misclassified as OII and GBM with a joint error rate of 23.3%.

4.2.3. Comparison of non-DL Classifiers Using Features Extracted with ResNet-18

We also evaluated the performance of 5 non-DL models on features extracted using the 18-layer residual network. To improve the ability of our non-DL classifiers to learn the features, it is necessary to carefully tune the model hyperparameters. For this reason, we apply Bayesian optimization to automatically tune the hyperparameters of our ML models. This optimization technique is a robust technique that models the objective function used to train an non-DL classifier as a Gaussian process [89]. The goal is to find the optimum model parameters that minimize an objective function on some bounded set.



(a) (b)

(c)

Figure 4.7: (a) The estimated 2-dimensional objective function versus the non-DL model hyperparameters was obtained using the Bayesian optimization technique for Decision tree Classifier. (b) The estimated 3-dimensional objective function model for Naïve Bayes Classifier. (c) Hyperparameter optimization for the kNN classifier showing the effects of the distance metric (Distance) and the number of nearest neighbors (NumNeighbors).

The values of the critical model hyperparameters for the different classifiers obtained from the Bayesian Optimization using the trained data are given below

- Naïve Bayes: Distribution Name = Kernel, Width = 0.23137
- Decision Tree: MinLeafSize = 63
- ANN: Activations = relu, Standard Size = True, Lambda = $1.2169e^{-6}$, Layer Size = 66
- KNN: Number of Neighbors = 1, Distance Metric (Distance) = Cosine
- SVM Multiclass: Coding = OneVsOne, Box Constraint = 0.0011045, Kernel scale = 0.17913

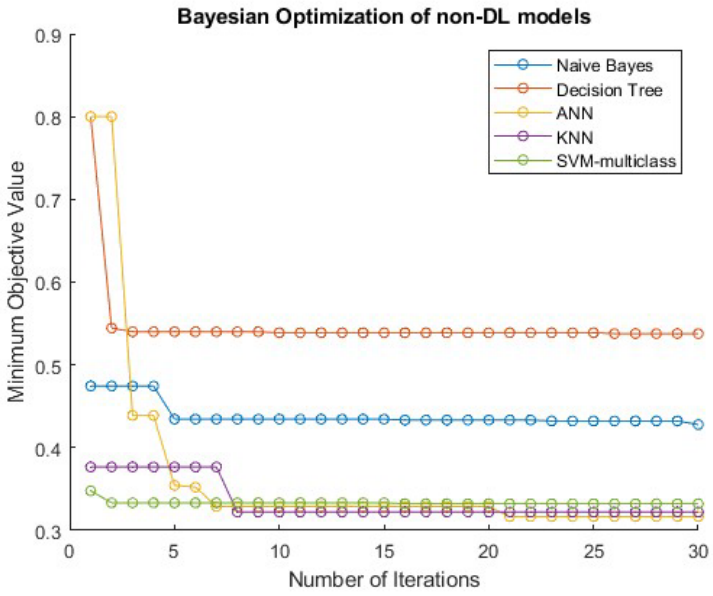


Figure 4.8: Minimum objective function plot for the 5 non-DL classifiers.

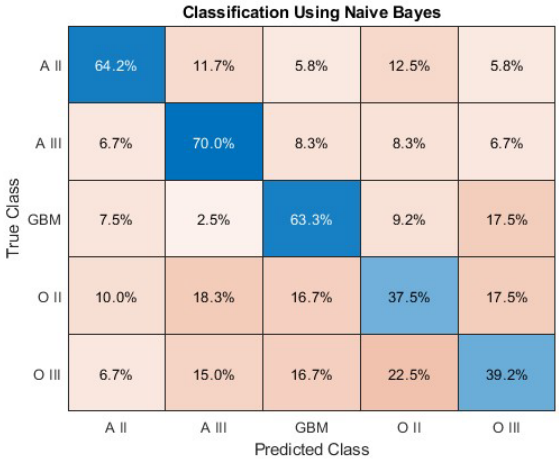
Fig. 4.7 shows that artificial neural network (ANN) has the lowest minimum objective value followed by KNN and the SVM multiclass. The decision tree classifier had the highest minimum objective value. This minimum objective function plot reflects the performances of our non-DL classifiers. The high minimum objective value for the decision tree classifier indicates that the decision tree classifier did not perform well on our dataset. On the other hand, the low minimum

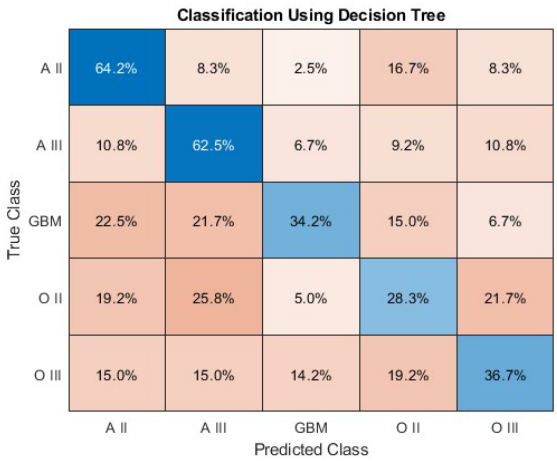
objective value for the ANN indicates that it had the best performance of all non-DL classifier for our dataset.

Table 4.6: Accuracy for each non-DL model

Non-DL Model	Accuracy (%)
Naïve Bayes	65.8
Decision Trees	54.2
ANN	83.8
KNN	83.6
SVM Multiclass	82

ANN achived the best accuracy with 83.8% followed by KNN with 83.6% and SVM with 82%. Decision tree had the worst accuracy with 54.2%.





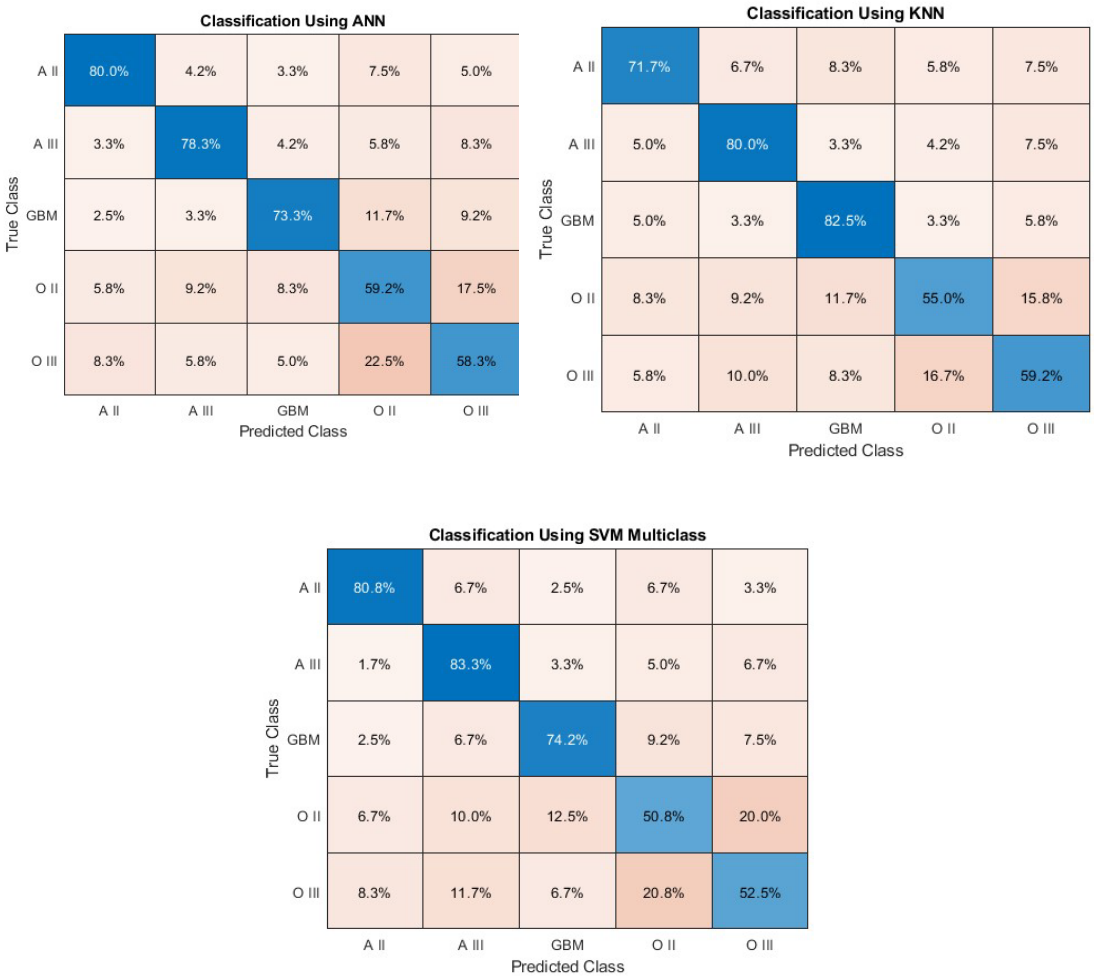


Figure 4.9: The confusion matrices of the best and worst non-DL model

The summary of the confusion matrices for each non-DL classifier is shown below

Table 4.7: Summary of the non-DL confusion natrices

Non-DL Models	**Best Classified class**	**Worst Classified Class**	**Most Misclassified Class**
Naïve Bayes	AIII	OII	OII
Decision Tree	AII	OII	AIII
ANN	AII	OIII	OII
KNN	GBM	OII	OII
SVM Multiclass	AIII	OII	OII

AIII had the highest classification rate with an average rate of 75% followed by AII with 72%. OII had the highest misclassification rate followed by OIII. OII had an average misclassification rate of 20% with OIII while the vice versa had an average error rate of 19%.

4.3 Comparing the Accuracy of All Models

We evaluated varieties of DL and non-DL models. DL models did not perform better than some of the non-DL models supporting the reason why DL models require high volume of data. However, in overview, the DL models did perform better than the non-DL models, thus, emphasizing the significance of the ability of DL models to learn high-level features from data in an incremental manner. This makes DL models advantageous over non-DL models when analyzing high dimensional data.

Table 4.8: Table Showing the Accuracy of All Models

Models	**Accuracy (%)**
GoogLeNet	76
ResNet-50	79
Inception-V3	80
DenseNet-201	82
Xception	76
Ensemble (Bag)	40.2
Naïve Bayes	65.8
Decision Tree	54.2
ANN	83.8
KNN	83.6
SVM Multiclass	82

Table 4.8 shows the accuracy of all models employed in this study. ANN proves to be the best model based on their accuracies in this study. However, the DL models achieved higher classification rates for all classes in terms of precision, clearly seen from their confusion matrices. Additionally, the decision tree classifier and the ensemble method both performed poorly, indicating that our dataset may not do so well with tree-like models.

5 CONCLUSION

A monoclonal antibody specific for IDH1 RI32H mutation known to be efficient in detecting single infiltrating tumor cells was administered for this study. Data augmentation was employed to increase the volume of our training set and enhance the accuracy of the disease classification task. To classify brain tumors of different classes, a variety of transfer learning DL and non-DL approaches were evaluated in this study.

For the original number of classes, DenseNet-201 achieved the best accuracy at 80%. To improve the accuracy of each model, we reduced the original classes from 8 classes to 5 classes. We discovered that this proposed method enhanced the accuracy and precision of each model. DenseNet-201 also achieved the best accuracy at 82% among the DL models for the reduced classes and seemed to generalize much better.

Ensemble method (Bag) with histogram of oriented gradients (Bag + HOG) and a variety of non-DL models with features extracted from ResNet-18 (non-DL + CNN) are two feature-based ML classification techniques also evaluated. The former achieved an accuracy of 40.2%. For the latter, ANN achieved the best accuracy at 83.8% followed by KNN with 83.6% and then SVM with 82%.

The astrocytoma grade 3 and grade 2 tumors were the best-classified classes. This could be attributed to the fact that grade 3 astrocytoma otherwise known as anaplastic astrocytoma, displays a higher degree of cellular abnormalities, and evidence of mitosis (proliferation) in comparison to grade 2 astrocytoma. On the other hand, the oligodendroglioma grade 2 was often misclassified as oligodendroglioma grade 3 and vice versa. According to [88], calcification, and the cortical-subcortical location, most often found in the frontal lobe, are regarded as characteristic features of oligodendrogliomas. Consequently, minimal to moderate enhancement

and averagely increased perfusions are common in oligodendrogliomas making differentiation of grade 2 and grade 3 oligodendroglioma challenging. Furthermore, edema, hemorrhage, cystic degeneration, and contrast enhancement are mostly seen in oligodendroglioma grade 3, however, they have been known to be also present in grade 2 oligodendrogliomas. A machine-learning model based on radiomics for efficient detection of grade 2 and grade 3 oligodendrogliomas was proposed in their study.

This study shows promising insights in detecting and diagnosing different glioma grades but will require further investigation. We could try using denoising and regularization techniques to evaluate the performance of each model. We could also explore other feature extraction techniques, such as wavelet scattering, which has shown promising results in recent medical image studies. However, increasing the volume of our dataset will undoubtedly yield better results.

5.1 Limitation of Study

The limitations of the study include the insufficient volume of available data. Due to their inherent complexity, and a structure of many layers, deep learning models need a large amount of data to perform effectively. Hence, the DL models applied in this study could not outperform several non-DL classifiers like the ANN, KNN and the SVM when considering their classification accuracies. Another factor is the complexity of the feature representation of the brain tumor histopathological images. The number of data points required for good performance of our non-DL classifiers increases exponentially as the feature dimensionality increases. The reason is that we would need more data points for any given combination of features, for any non-DL model to be valid. Lastly, the lack of clear boundaries between gliomas makes distinguishing each class difficult. Sometimes a biopsy is needed to make a proper diagnosis.

5.2 Implication of this Study

The implications of this study are outlined below

1. The use of deep learning in this study allowed us to extract and learn feature representation automatically from images characterized by data complexity. This indicates that deep learning is a promising tool for assisting pathologists in making prognoses and providing better treatment for patients.
2. This study also shows that there are potential insights that could be derived when cell tissues are stained with antibodies that target specific biomarkers.
3. This study also underscores the importance of the evolution of artificial intelligence in eradicating the subjective process of pathologists in analyzing each digital slide.

REFERENCES

1. Hamida, A. B., Devanne, M., Weber, J., Truntzer, C., Derangère, V., Ghiringhelli, F., ... & Wemmert, C. (2021). Deep learning for colon cancer histopathological images analysis. *Computers in Biology and Medicine*, *136*, 104730.
2. Ayyad, S. M., Shehata, M., Shalaby, A., Abou El-Ghar, M., Ghazal, M., El-Melegy, M., ... & El-Baz, A. (2021). Role of AI and histopathological images in detecting prostate cancer: a survey. *Sensors*, *21*(8), 2586.
3. Xie, J., Liu, R., Luttrell IV, J., & Zhang, C. (2019). Deep learning based analysis of histopathological images of breast cancer. *Frontiers in genetics*, *10*, 80.
4. Suvarna, K. S., Layton, C., & Bancroft, J. D. (Eds.). (2018). *Bancroft's theory and practice of histological techniques E-Book*. Elsevier health sciences.
5. Banerji, S., & Mitra, S. (2022). Deep learning in histopathology: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(1), e1439.
6. Belsare, A. D., & Mushrif, M. M. (2012). Histopathological image analysis using image processing techniques: An overview. *Signal & Image Processing*, *3*(4), 23.
7. Hameed, Z., Zahia, S., Garcia-Zapirain, B., Javier Aguirre, J., & María Vanegas, A. (2020). Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors*, *20*(16), 4373.
8. Cai, L., Gao, J., & Zhao, D. (2020). A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine*, *8*(11).

9. Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, *65*, 101759.
10. Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, *148*(3), 574.
11. Smeda, K. (2019, October 31). Understand the architecture of CNN. *Towards Data Science.* https://towardsdatascience.com/understand-the-architecture-of-cnn-90a25e244c7
12. Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Conceptual understanding of convolutional neural network-a deep learning approach. *Procedia computer science*, *132*, 679-688.
13. Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, *29*(2), 102-127.
14. Eaton-Rosen, Z., Bragman, F., Ourselin, S., & Cardoso, M. J. (2018). Improving data augmentation for medical image segmentation.
15. Nalepa, J., Marcinkiewicz, M., & Kawulok, M. (2019). Data augmentation for brain-tumor segmentation: a review. *Frontiers in computational neuroscience*, *13*, 83.
16. Sanya, S. (2021, June 14). Offline Data Augmentation for multiple images. *Analytics Vidhya*. https://www.analyticsvidhya.com/blog/2021/06/offline-data-augmentation-for-multiple-images/
17. Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), 143-150.

18. Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359.
19. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*(1), 43-76.
20. Accessed:2022-07-01. [Online]. Available: https://www.mathworks.com/help/stats/bayesian-optimization-algorithm.html
21. Kaushal, C., Bhat, S., Koundal, D., & Singla, A. (2019). Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological images. *Irbm*, *40*(4), 211-227.
22. Khalid, S., Khalil, T., & Nasreen, S. (2014, August). A survey of feature selection and feature extraction techniques in machine learning. In *2014 science and information conference* (pp. 372-378). IEEE.
23. Zheng, Y., Jiang, Z., Xie, F., Zhang, H., Ma, Y., Shi, H., & Zhao, Y. (2017). Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification. *Pattern Recognition*, *71*, 14-25.
24. Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee.
25. Ladha, L., & Deepa, T. (2011). Feature selection methods and algorithms. *International journal on computer science and engineering*, *3*(5), 1787-1797.
26. Kowal, M., Filipczuk, P., Obuchowicz, A., Korbicz, J., & Monczak, R. (2013). Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Computers in biology and medicine*, *43*(10), 1563-1572.

27. Osborne, J. D., Gao, S., Chen, W. B., Andea, A., & Zhang, C. (2011, March). Machine classification of melanoma and nevi from skin lesions. In *Proceedings of the 2011 ACM Symposium on Applied Computing* (pp. 100-105).
28. Olgun, G., Sokmensuer, C., & Gunduz-Demir, C. (2013). Local object patterns for the representation and classification of colon tissue images. *IEEE journal of biomedical and health informatics*, *18*(4), 1390-1396.
29. Krishnan, M. M. R., Pal, M., Bomminayuni, S. K., Chakraborty, C., Paul, R. R., Chatterjee, J., & Ray, A. K. (2009). Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis—An SVM based approach. *Computers in biology and medicine*, *39*(12), 1096-1104.
30. Dundar, M. M., Badve, S., Bilgin, G., Raykar, V., Jain, R., Sertel, O., & Gurcan, M. N. (2011). Computerized classification of intraductal breast lesions using histopathological images. *IEEE Transactions on Biomedical Engineering*, *58*(7), 1977-1984.
31. Filipczuk, P., Fevens, T., Krzyżak, A., & Monczak, R. (2013). Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. *IEEE transactions on medical imaging*, *32*(12), 2169-2178.
32. Zhang, Y., Zhang, B., Coenen, F., & Lu, W. (2013). Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles. *Machine vision and applications*, *24*(7), 1405-1420.
33. Mazo, C., Alegre, E., & Trujillo, M. (2017). Classification of cardiovascular tissues using LBP based descriptors and a cascade SVM. *Computer methods and programs in biomedicine*, *147*, 1-10.

34. Chan, A., & Tuszynski, J. A. (2016). Automatic prediction of tumour malignancy in breast cancer with fractal dimension. *Royal Society open science*, *3*(12), 160558.

35. Harai, Y., & Tanaka, T. (2015, November). Automatic Diagnosis Support System Using Nuclear and Luminal Features. In *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-8). IEEE.

36. Zhang, R., Shen, J., Wei, F., Li, X., & Sangaiah, A. K. (2017). Medical image classification based on multi-scale non-negative sparse coding. *Artificial intelligence in medicine*, *83*, 44-51.

37. Atupelage, C., Nagahashi, H., Yamaguchi, M., Abe, T., Hashiguchi, A., & Sakamoto, M. (2013). Computational grading of hepatocellular carcinoma using multifractal feature description. *Computerized Medical Imaging and Graphics*, *37*(1), 61-71.

38. Vanderbeck, S., Bockhorst, J., Komorowski, R., Kleiner, D. E., & Gawrieh, S. (2014). Automatic classification of white regions in liver biopsies by supervised machine learning. *Human pathology*, *45*(4), 785-792.

39. Rahman, T. Y., Mahanta, L. B., Chakraborty, C., Das, A. K., & Sarma, J. D. (2018). Textural pattern classification for oral squamous cell carcinoma. *Journal of microscopy*, *269*(1), 85-93.

40. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016, July). Breast cancer histopathological image classification using convolutional neural networks. In *2016 international joint conference on neural networks (IJCNN)* (pp. 2560-2567). IEEE.

41. Orlov, N. V., Chen, W. W., Eckley, D. M., Macura, T. J., Shamir, L., Jaffe, E. S., & Goldberg, I. G. (2010). Automatic classification of lymphoma images with transform-based

global features. *IEEE Transactions on Information Technology in Biomedicine*, *14*(4), 1003-1013.

42. Kong, J., Sertel, O., Shimada, H., Boyer, K. L., Saltz, J. H., & Gurcan, M. N. (2009). Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation. *Pattern Recognition*, *42*(6), 1080-1092.
43. Zarella, M. D., Breen, D. E., Reza, A., Milutinovic, A., & Garcia, F. U. (2015). Lymph Node Metastasis Status in Breast Carcinoma Can Be Predicted via Image Analysis of Tumor Histology. *Analytical and quantitative cytopathology and histopathology*, *37*(5), 273-285.
44. DiFranco, M. D., Reynolds, H. M., Mitchell, C., Williams, S., Allan, P., & Haworth, A. (2015, April). Performance assessment of automated tissue characterization for prostate H and E stained histopathology. In *Medical Imaging 2015: Digital Pathology* (Vol. 9420, pp. 143-151). SPIE.
45. Kruk, M., Kurek, J., Osowski, S., Koktysz, R., Swiderski, B., & Markiewicz, T. (2017). Ensemble of classifiers and wavelet transformation for improved recognition of Fuhrman grading in clear-cell renal carcinoma. *Biocybernetics and Biomedical Engineering*, *37*(3), 357-364.
46. Valkonen, M., Kartasalo, K., Liimatainen, K., Nykter, M., Latonen, L., & Ruusuvuori, P. (2017). Metastasis detection from whole slide images using local features and random forests. *Cytometry Part A*, *91*(6), 555-565.
47. Gertych, A., Ing, N., Ma, Z., Fuchs, T. J., Salman, S., Mohanty, S., ... & Knudsen, B. S. (2015). Machine learning approaches to analyze histological images of tissues from radical prostatectomies. *Computerized Medical Imaging and Graphics*, *46*, 197-208.

48. Romo-Bucheli, D., Corredor, G., García-Arteaga, J. D., Arias, V., & Romero, E. (2017, January). Nuclei graph local features for basal cell carcinoma classification in whole slide images. In *12th International Symposium on Medical Information Processing and Analysis* (Vol. 10160, pp. 205-213). SPIE.
49. Malon, C., Brachtel, E., Cosatto, E., Graf, H. P., Kurata, A., Kuroda, M., ... & Yagi, Y. (2012). Mitotic figure recognition: Agreement among pathologists and computerized detector. *Analytical Cellular Pathology*, *35*(2), 97-100.
50. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.
51. Das, D. K., Bose, S., Maiti, A. K., Mitra, B., Mukherjee, G., & Dutta, P. K. (2018). Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. *Tissue and Cell*, *53*, 111-119.
52. Sharma, H., Zerbe, N., Klempert, I., Hellwich, O., & Hufnagl, P. (2017). Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Computerized Medical Imaging and Graphics*, *61*, 2-13.
53. Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O., & Hajirasouliha, I. (2018). Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine*, *27*, 317-328.
54. Zerhouni, E., Lányi, D., Viana, M., & Gabrani, M. (2017, April). Wide residual networks for mitosis detection. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 924-928). IEEE.

55. Kainz, P., Pfeiffer, M., & Urschler, M. (2017). Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. *PeerJ*, *5*, e3874.

56. Stanitsas, P., Cherian, A., Li, X., Truskinovsky, A., Morellas, V., & Papanikolopoulos, N. (2016, December). Evaluation of feature descriptors for cancerous tissue recognition. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 1490-1495). IEEE.

57. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016, July). Breast cancer histopathological image classification using convolutional neural networks. In *2016 international joint conference on neural networks (IJCNN)* (pp. 2560-2567). IEEE.

58. Li, W., Manivannan, S., Akbar, S., Zhang, J., Trucco, E., & McKenna, S. J. (2016, April). Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)* (pp. 1405-1408). IEEE.

59. Talo, M. (2019). Automated classification of histopathology images using transfer learning. *Artificial intelligence in medicine*, *101*, 101743.

60. Kwak, J. T., & Hewitt, S. M. (2017). Nuclear architecture analysis of prostate cancer via convolutional neural networks. *IEEE access*, *5*, 18526-18533.

61. Budak, Ü., Cömert, Z., Rashid, Z. N., Şengür, A., & Çıbuk, M. (2019). Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images. *Applied Soft Computing*, *85*, 105765.

62. Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific reports*, *7*(1), 1-10.

63. Nawaz, M., Sewissy, A. A., & Soliman, T. H. A. (2018). Multi-class breast cancer classification using deep learning convolutional neural network. *Int. J. Adv. Comput. Sci. Appl*, *9*(6), 316-332.

64. Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013, September). Mitosis detection in breast cancer histology images with deep neural networks. In *International conference on medical image computing and computer-assisted intervention* (pp. 411-418). Springer, Berlin, Heidelberg.

65. Sirinukunwattana, K., Raza, S. E. A., Tsang, Y. W., Snead, D. R., Cree, I. A., & Rajpoot, N. M. (2016). Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging*, *35*(5), 1196-1206.

66. Xu, J., Xiang, L., Liu, Q., Gilmore, H., Wu, J., Tang, J., & Madabhushi, A. (2015). Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE transactions on medical imaging*, *35*(1), 119-130.

67. Yang, X., Yeo, S. Y., Hong, J. M., Wong, S. T., Tang, W. T., Wu, Z. Z., ... & Su, Y. (2016). A deep learning approach for tumor tissue image classification. *IASTED Biomedical Engineering*.

68. Wang, H., Roa, A. C., Basavanhally, A. N., Gilmore, H. L., Shih, N., Feldman, M., ... & Madabhushi, A. (2014). Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, *1*(3), 034003.

69. Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., & Navab, N. (2016). Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE transactions on medical imaging*, *35*(5), 1313-1321.

70. Guo, Z., Liu, H., Ni, H., Wang, X., Su, M., Guo, W., ... & Qian, Y. (2019). A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. *Scientific reports*, *9*(1), 1-10.

71. Yan, R., Ren, F., Wang, Z., Wang, L., Zhang, T., Liu, Y., ... & Zhang, F. (2020). Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, *173*, 52-60.

72. De Matos, J., Britto, A. D. S., Oliveira, L. E., & Koerich, A. L. (2019, July). Double transfer learning for breast cancer histopathologic image classification. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

73. Vizcarra, J., Place, R., Tong, L., Gutman, D., & Wang, M. D. (2019, September). Fusion in breast cancer histology classification. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (pp. 485-493).

74. Brancati, N., De Pietro, G., Frucci, M., & Riccio, D. (2019). A deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images. *IEEE Access*, *7*, 44709-44720.

75. Ataky, S. T. M., De Matos, J., Britto, A. D. S., Oliveira, L. E., & Koerich, A. L. (2020, July). Data augmentation for histopathological images based on gaussian-laplacian pyramid blending. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

76. de Matos, J., de Souza Britto, A., de Oliveira, L. E., & Koerich, A. L. (2019, June). Texture CNN for histopathological image classification. In *2019 IEEE 32nd international symposium on computer-based medical systems (CBMS)* (pp. 580-583). IEEE.

77. Ciompi, F., Geessink, O., Bejnordi, B. E., De Souza, G. S., Baidoshvili, A., Litjens, G., ... & Van Der Laak, J. (2017, April). The importance of stain normalization in colorectal tissue classification with convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 160-163). IEEE.

78. Sheikh, T. S., Lee, Y., & Cho, M. (2020). Histopathological classification of breast cancer images using a multi-scale input and multi-feature network. *Cancers*, *12*(8), 2031.

79. Capper, D., Zentgraf, H., Balss, J., Hartmann, C., & Von Deimling, A. (2009). Monoclonal antibody is specific for IDH1 R132H mutation. *Acta neuropathologica*, *118*(5), 599-601.

80. Weng L. (2017). Object Detection for Dummies Part 1: Gradient Vector, HOG, and SS. *lilianweng.github.io.*

https://lilianweng.github.io/posts/2017-10-29-object-recognition-part-1/

81. *Tyagi, M. (2021, July 4). HOG (Histogram of Oriented Gradients): An Overview. Towards Data Science.*

https://towardsdatascience.com/hog-histogram-of-oriented-gradients-67ecd887675f

82. Tomasi, C. (2012). Histograms of oriented gradients. *Computer Vision Sampler*, 1-6.

83. Accessed:2022-07-01. [Online]. Available: https://www.mathworks.com/help/stats/bayesian-optimization-algorithm.html

84. Brownlee, J. (2019, October 9). How to implement Bayesian Optimization from Scratch in Python. *Machine Learning Mastery*.

https://machinelearningmastery.com/what-is-bayesian-optimization/

85. Wang, W. (2020, March 18). Bayesian Optimization Concept Explained in Layman Terms. *Towards Data Science.* https://towardsdatascience.com/bayesian-optimization-concept-explained-in-layman-terms-1d2bcdeaf12f

86. Capper, D., Zentgraf, H., Balss, J., Hartmann, C., & von Deimling, A. (2009). Monoclonal antibody specific for IDH1 R132H mutation. *Acta neuropathologica*, *118*(5), 599–601. https://doi.org/10.1007/s00401-009-0595-z

87. Zhao, S. S., Feng, X. L., Hu, Y. C., Han, Y., Tian, Q., Sun, Y. Z., ... & Wang, W. (2020). Better efficacy in differentiating WHO grade II from III oligodendrogliomas with machine-learning than radiologist's reading from conventional T1 contrast-enhanced and fluid attenuated inversion recovery images. *BMC neurology*, *20*(1), 1-10.

88. Pierpaolo, P., Vikram C. P. Astrocytoma Tumors. *American Association of Neurological Surgeons*. https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Astrocytoma-Tumors

89. Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*.