Enhancing Histopathological Tissue Accuracy Using: OPCNN And BERT

1st Swathi S
Computer Science And Engineering
College of Engineering, Perumon
Kollam, India

2nd Dr . K. S. Angel Viji
Computer Science And Engineering
College of Engineering, Perumon
Kollam, India

***Abstract*—Histopathological tissues are not only critical to cancer diagnosis, but they also provide useful tumor microenvironment information for cancer research. Current CNN classification has already shown strong feature representation ability and promising outcomes for histopathology tissue classification. In this paper, we propose a method using optimized convolutional neural networks (OPCNN) and Bidirectional Encoder Representations from Transformers (BERT). The convolutional Auto encoder's aim is to learn an input function to reconstruct the input to an output of fewer dimensions. Tissue Classification is compelled to learn numerical changes that carry the most useful details about the structure of data in order for the deciphering part to operate well in the rebuilding task. The BERT model's remarkable performance could possibly be attributable to the fact that it is bidirectionally trained. This implies that BERT, which is built on the Transformer model architecture, uses its self-attention mechanism during training to learn information from both the left and right sides, resulting in a deep understanding of the context. On two downstream tasks, picture classification, and semantic segmentation, we fine-tune the pre-trained BERT and self-supervised learning. The output of the BERT layer is routed into OPCNN, which then passes the output to a completely linked bulky layer, which produces a single posture as its final output. On the Lung Colon Cancer Histopathological Image Dataset, we subjected the proposed approach to the test. The findings from the study indicate that the proposed technique can improve tissue-level accuracy for classification by up to 96.91% over time. It significantly shortens the processing time.**

***Index Terms*—OPCNN, BERT, Histopathological tissue classification.**

I. Introduction

Histopathology is the microscopic inspection of tissue to investigate disease manifestations[1][2]. In clinical medicine, histopathology refers to a pathologist's study of a biopsy or surgical specimen after it has been processed and histological sections have been placed onto glass slides. Histopathology depends on tissue samples obtained through endoscopy, colonoscopy, and colposcopy, as well as surgical procedures such as a breast biopsy. Histopathology examinations are used to identify the following diseases: Colitis ulcerative, Crohn's disease is a type of IBD, Fibroids in the uterus, Cancer, Infections. Cancer has emerged as a major public health issue. According to data from the IARC (International Agency for Research on Cancer) of the WHO (World Health Organisation) and the GBD (Global Burden of Disease Cancer Collaboration), cancer cases increased by 28% between 2006 and 2016, and there will be an additional 2.7 million cases of cancer worldwide in 2030. Automated histopathological image analysis and classification have recently emerged as a major research issue in medical imaging, with a growing need for developing quantitative image analysis methods to supplement pathologists' efforts in the diagnosis process. As a result, an emerging class of medical imaging problems concentrates on the creation of computerized frameworks for classifying histopathological images [1]-[5]. These advanced image analysis procedures were developed with three main goals in mind: (i) releasing pathologists' job stress by separating out evidently diseased and healthy cases, facilitating specialists to devote more time to more complex cases; (ii) cutting inter-expert fluctuation; and (iii) recognizing the root causes for particular diagnoses that physicians may be unaware of.

Convolutional neural networks (CNN) have proven exceptional performance in image classification problems [25], with a variety of sorting foundations including The ResNet protocol [27], ShuffLeNet [31], EfficientNet [33], and others. They were quickly expanded to include medical image classification [26], [17], and histopathological image classification [20, 29]. Han et al. [15] demonstrated a CNN-based multiclassification approach to histopathological cell categorization in cancer of the breast that achieves greater than 94% patch-level precision under four magnifying lens factors. Tsai and Tao [14] We tested five common classification backbones on colorectal tissue classification using 100,000 training image patches. All of the backbones are 95% accurate. Previous research showed that current neural network backbones have a solid depiction of features ability and have produced promising outcomes for histopathology material grouping.

Here propose a method using optimized convolutional neural networks (OPCNN) and Bidirectional Encoder Representations from Transformers (BERT). We look at how to make more productive use of the features extracted from existing CNN approaches to boost the performance of classification while also increasing model generalizability, adaptability, and robustness in this paper.

This paper's offerings have been outlined below:

- During the pre-training step, autoencoding-type recon-

Identify applicable funding agency here. If none, delete this.

struction, i.e. image tokenizer, is used to tokenize an image into discrete visual tokens.

- We randomly mask some image grey areas and substitute them with a special mask embedding. The changes are then routed to a backbone vision Transformer. The pre-training process guesses the visual tokens of the original images using the image encoding vectors. In addition, we randomly mask some image patches and send the corrupted input to Transformer. Instead of raw pixels from masked patches, the model learns to recover the visual tokens of the original picture.
- The BERT layer's production flows into OPCNN, which is then sent to a totally linked brittle stage, producing an identical stride as the final result.

II. RELATED WORKS

A. *Histopathological Image Classification*

Jiatai Lin [6] proposed using PDBL to improve histopathological tissue classification. Pyramidal Deep-Broad Learning (PDBL) is a simple built-to-be module that works with any CNN categorization core. For each input picture, an image pyramid is built to extract the layered contextual information. In this method, a Deep-Broad block is suggested for each level of the structure in order to find all of the multiple-scale deep-broad characteristics collected by The CNN algorithm vertebrae from the bottom to the top. With the development of CNN models, most histopathological image categorization prototypes originate from popular classification networks that originate from organic image classification. Histopathological sorting of images, on the other hand, faces a number of difficulties, including high resolution of images, label absence, and multiple habitats incorporation of data [26]. The WSI-Net simulation has been recommended to add an extra classification branch to discard the usual tissue in order to save computational assets.

In this paper, we propose a method using optimized convolutional neural networks (OPCNN) and Bidirectional Encoder Representations from Transformers (BERT).

B. *Image preprocessing*

Image pre-processing refers to operations performed on images at the most basic level of realism. If the entropy is an evaluation for data, then these actions lower instead of increasing the image's data content. Preparation is used to better image data through the elimination of imperfections or increasing specific optical characteristics that are essential to following data analysis and processing tasks. Although geometric image transformations are classified as pre-processing procedures because similar techniques are used, the goal of pre-processing is an enhancement to the image information that removes undesirable distortions or enhances some image elements that are useful for the next stage of processing.

C. *Principal Component Analysis (PCA)*

Principal Component Analysis reduces the number of features while retaining as much information as possible. Begin by subtracting the mean from each data point to normalize the variables. It is critical to normalize the predictor because the original predictors may be on a different scale and contribute considerably to variance. Next, compute the data's covariance matrix, which measures how two variables move together. These elements are listed in decreasing order of significance. A matrix of covariance is a square matrix that expresses the variance of sample components as well as the covariance between two datasets. Variance is a gauge of dispersion and is described as the spread of data from the given dataset's mean. The covariance of two variables is calculated and used to determine how the variables vary together.

D. *Feature Extraction and Feature Mapping*

Dimensionality reduction in image classification can be performed in two ways: feature extraction and feature selection. In feature selection methods, the most significant and pertinent features remain intact while the rest of the characteristics get eliminated. Chi-Square Statistics is one of the most efficient feature selection algorithms. Its purpose is to determine the relationships between variables that are categorical. It is used to examine a and b's lack of sovereignty, as well as to compare extremeness to the distribution of chi-squares with a single attribute of freedom. Chi-square is used for two types of evaluations: independence tests and goodness of fit tests. A test for independence is utilized for feature selection, and the goal label's dependence on the feature is evaluated. The feature correlation is investigated using Chi-square. The features with an associated pattern are kept, while the rest are removed.

E. *BERT*

BERT is an acronym that stands for Bidirectional Encoder Representations from Transformer. Traditional deep neural network models are outperformed by the proposed technique. In order to analyze the link, four data models are developed. In the first model, all characteristics are used for categorization without any preprocessing. In the second model, the non-reduced feature accumulation is employed after preprocessing. The third and fourth models are created via reducing dimensionality. The embedding layer is the model's initial layer, and it translates input data and article bodies to a size of 70x70. There are a total of 5000 images (5000 for each category). The weights generated from matrix multiplication will be used to create a vector for each image in the output matrix. These vectors are used by the BERT layer to determine pertinent data. The output of the BERT layer is funneled into OPCNN to the end result.

BERT is an RNN-based design that is used in deep learning. This is where we compute the Memory units and build the models. We offer threshold numbers ranging from 0 to 1. Following the creation of the model, we construct three layers of memory units and calculate their maximum value. The memory units are used in the computation.

F. *Deep Learning Model Generation*

The purpose of a convolutional auto-encoder is to learn an input function in order to rebuild the input into a smaller-

dimensional output. When we throw the model's output, we can use the activation function to help us conduct classification or regression tasks. Before making any predictions, the activation function is called after the hidden layer to learn a non-linear type of linear mapping.

The TensorFlow deep learning platform was used for developing our sorting method. Over the learning procedure, the algorithm known as SGD was used to optimize the histopathological picture collection of lung and colon cancer by iterating through 20 epochs. The exponential decrement methodology is then used to minimize the speed of learning and ensure that the system cycles quickly during the very first training stage. we learn a total of 32 filters. Max pooling is then used to reduce the spatial dimensions of the output volume. The pooling layer employs the max procedure to reduce the features in the feature map. Choosing the highest number allows you to capture the most important features while minimizing the amount of computation needed in the next layer. We then learn 64 filters. To minimize the spatial dimensions, max pooling is utilized once more. In the initial layers, I suggest commencing with barriers in the [32, 64, 128] range and growing to [256, 512, 1024] in the more advanced layers of stuff. This also provides to greater stability in the future, making it easier to discover the best solution.Fig.1 exhibits the workflow of the proposed model.

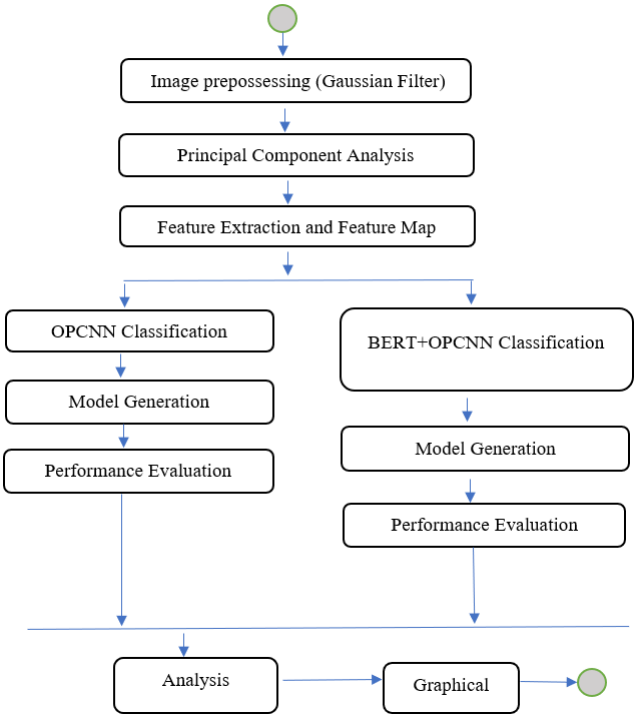

Fig. 1. Workflow of the proposed method

III. Implementation and Training Details

A. Experimental Setup

TThe dataset is separated into three distinct sections: training, validation, and testing, with a ratio of 80%, 10%, and 10%, accordingly. Each of the tests happened on a Google Colab notebook powered by an AMD Ryzen 5, 5500U with Radeon Graphics 2.10 GHz processor. The OPCNN-BERT model is trained and tested using Keras with the TensorFlow backend. The coding language Python and multiple typical collections such as pandas, NumPy, Matplotlib, Seaborn, TensorFlow, Keras, and Scikit-learn are used to carry out the suggested framework. The simulations took out on an aggregate of 20 epochs with a total number of batches of 1024 and an average rate of learning of 0.001.

B. LC25000 Dataset

[39] published a lung and colon histopathological image dataset(LC25000 Dataset) to improve computer-aided automated analysis of tissues. The database contains 25,000 color photographs, which have been separated into five groups of 5,000 images each. All photos are 768 x 768 pixels in size and saved as jpegs. The dataset is available for distribution as a 1.85 GB zip file called LC25000.zip. Fig.2 illustrates LC25000 Dataset.

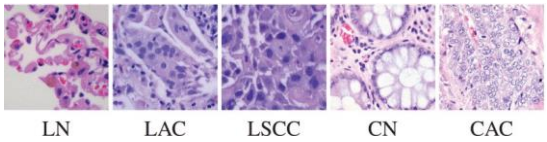

Fig. 2. LC25000

IV. Experimental Results

A. Loss and Accuracy

Loss and Accuracy are the important metrics to be taken care of in the training and validation phases. The loss of training measurement is used to evaluate how well a model created using deep learning fits the training data. That is, it determines the model's prediction error on its initial training set. The learning loss is computed after each batch by combining the sum of errors for each sample in the training set. A training loss sloped is frequently utilized for depicting this. Validation loss, on the other hand, is a statistic used to evaluate the achievement of a deep learning model on the data set being validated. The loss caused by validation is calculated by adding the errors for every value in the set of validation results and comparing it to the original training loss.

Fig.3 depicts the first phase of the proposed model's performance in the training and validation phases in terms of accuracy and loss using an extended version of OPCNN alone. The obtained training and validation accuracy over the first phase of the proposed system is 59.22 % and 48.0%, respectively. In the same way, training and validation loss is 0.9478 and 0.9613, respectively. TABLE 1 depicts the performance per epoch of the first phase of the proposed method.

The proposed method is implemented by using an extended version of OPCNN and BERT to achieve greater performance. In the model accuracy plot, the blue line represents validation

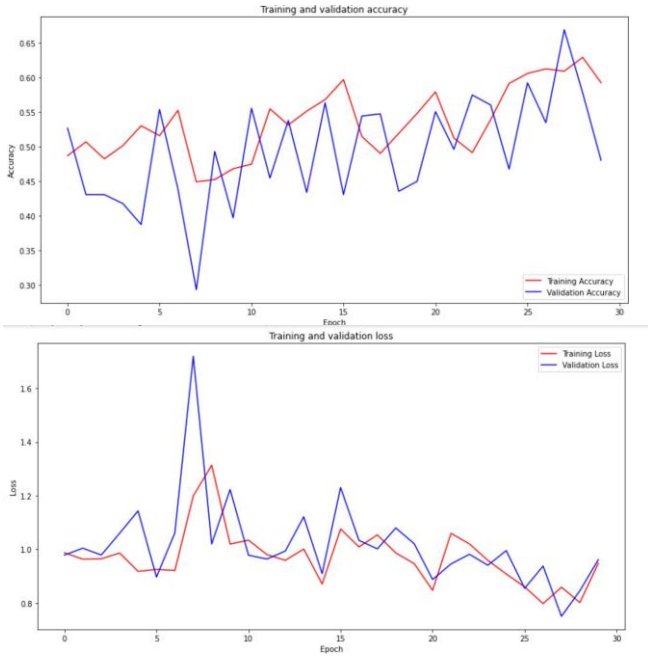

Fig. 3. Training accuracy and training loss

TABLE I
ABLATION OF OPCNN(FIRST PHASE) DESIGN (LOSS & ACCURACY)

Epoch	LC25000 Dataset			
	Loss	*Accuracy*	*Val Loss*	*Val Accuracy*
1	0.9869	0.4867	0.9782	0.5264
2	0.9630	0.5067	1.0039	0.4304
3	0.9643	0.4822	0.9783	0.4304
4	0.9643	0.4822	0.9783	0.4304
5	0.9175	0.5300	1.1431	0.3872
6	0.9249	0.5156	0.8963	0.5536
7	0.9207	0.5522	1.0613	0.4384
8	1.1988	0.4489	1.7194	0.2928
9	1.3132	0.4522	1.0196	0.4928
10	1.0193	0.4678	1.2225	0.3968
11	1.0340	0.4744	0.9778	0.5552
12	0.9798	0.5544	0.9634	0.4544
13	0.9588	0.5311	0.9934	0.5376
14	1.0007	0.5511	1.1210	0.4336
15	0.8702	0.5678	0.9090	0.5632
16	1.0754	0.5967	1.2303	0.4304
17	1.0087	0.5144	1.0331	0.5440
18	1.0536	0.4900	1.0013	0.5472
19	0.9861	0.5189	1.0794	0.4352
20	0.9463	0.5478	1.0207	0.4496
21	0.8470	0.5789	0.8871	0.5504
22	1.0594	0.5122	0.9454	0.4960
23	1.0200	0.4911	0.9812	0.5744
24	0.9585	0.5389	0.9404	0.5600
25	0.9076	0.5911	0.9949	0.4672
26	0.8597	0.6056	0.8543	0.5920
27	0.7971	0.6122	0.9372	0.5344
28	0.8584	0.6089	0.7500	0.6688
29	0.8010	0.6289	0.8454	0.5776
30	0.9478	0.5922	0.9613	0.4800

efficiency, while the red line represents accuracy during training. Similarly, in the Model loss plot, the red line represents training loss, while the blue line represents validation loss. The proposed system's training and validation accuracies are 95.20% and 96.51%, respectively. Similarly, the training and validation losses are 0.0929 and 0.1005, respectively. TABLE II depicts the performance per epoch of the proposed method.

TABLE II
ABLATION OF OPCNN+BERT DESIGN (LOSS & ACCURACY)

Epoch	LC25000 Dataset			
	Loss	*Accuracy*	*Val Loss*	*Val Accuracy*
1	1.2949	0.3836	1.0017	0.4869
2	0.8929	0.6019	0.6265	0.7413
3	0.7178	0.6983	0.5702	0.7787
4	0.6338	0.7351	1.5474	0.3406
5	0.6338	0.7351	1.5474	0.3406
6	0.6338	0.7351	1.5474	0.3406
7	0.6338	0.7351	1.5474	0.3406
8	0.3416	0.8658	0.2676	0.8987
9	0.3117	0.8793	0.2896	0.8923
10	0.2765	0.8935	0.2320	0.9093
11	0.2389	0.9074	0.2373	0.9085
12	0.2232	0.9124	0.1733	0.9338
13	0.2055	0.9207	0.1701	0.9333
14	0.1912	0.9267	0.1553	0.9395
15	0.2049	0.9229	0.1620	0.9373
16	0.1605	0.9405	0.1369	0.9466
17	0.1502	0.9437	0.1196	0.9563
18	0.1388	0.9471	0.1272	0.9496
19	0.2190	0.9187	0.2043	0.9210
20	0.1308	0.9520	0.1005	0.9651

B. Performance Analysis

To assess the efficacy of our model with previous models in research, we repeated the same steps but with 25,000 images from the LC25000 file. compares the results of colon and lung cancer subtype categorization gathered with various methods using the exact same dataset. This study used the public dataset generated from LC25000 to investigate the efficacy of homology-based image processing in large datasets. The results of the general public are the set show that homology-based processing of images beats typical texture analysis in recognizing lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue. Figure 4 depicts the results of the performance analysis.

Method	No of epoch	Validation loss	Accuracy
CNN Model	30	0.5421	0.8902
OPCNN+BERT Model	20	0.0929	0.9691

Fig. 4. Performance analysis results

Our research has shown that by using characteristic engineering, we can achieve results comparable to Deep Learning approaches. BERT had a 97% accuracy rate in classifying

colon and lung cancer subtypes. Fig. 5 illustrates the comparison of the proposed model's Accuracy and loss with the existing system.

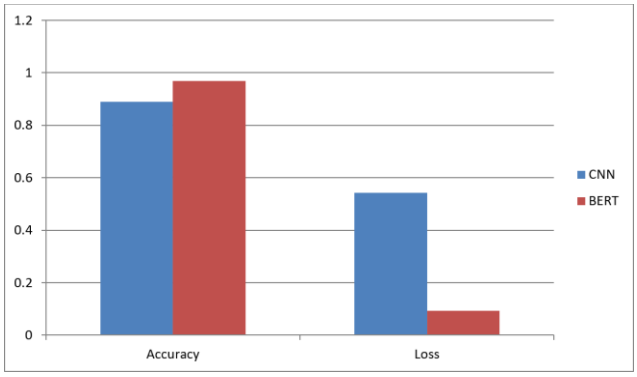

Fig. 5. Performance analysis of the proposed model with CNN

C. Accuracy, Precision, Recall, F1 Score

Precision, Recall (or sensitivity), and F1-score are used to measure and analyze the performance of the model using synthetic data augmentation technique.

The most basic and straightforward performance measure is accuracy, which is just a ratio of accurately anticipated observations to total observations. The model is excellent if we have great precision. Yes, accuracy is an excellent metric, but only when the amounts of false positives and false negatives are close to equal. As a result, assess the performance of your model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision can be described as the quantity of precisely anticipated positive observations to all imagined positive observations.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The ratio of precisely anticipated positive outcomes to all observations in the actual class is collectively referred to as recall.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F score is also called F1 score - F1 Score is the weighted average of Precision and Recall. It is used to evaluate binary classification systems, which classify examples as positive or negative.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

We can see that we have about approx 96.91% accuracy for the training set. Shown in Fig.4

Then the model will restore weights to their best validation performance. fig.6 illustrates the final result of the proposed method.

```
model.evaluate(xtrain,ytrain)

586/586 [==============================] - 95s 162ms/step - loss: 0.0929 - accuracy: 0.9691
[0.09292329847812653, 0.9691200256347656]
```

Fig. 6. Result of Proposed Model

V. CONCLUSION AND FUTURE SCOPE

Histopathological slides are not only important for cancer diagnosis, but they also provide essential tumor microenvironment information for cancer research. We repeated the equivalent methods but used 25,000 photos of colon and lung cancer from the LC25000 collection to assess our model with existing models in the available research. The paper compares the results of gastrointestinal and lung tumor subtype classification using various approaches on the same set.

Our research has shown that by using feature engineering, we may achieve outcomes comparable to Deep Learning techniques. BERT has a 97% accuracy rate in classifying cancer subtypes. We can see that the model for each organ type is more efficient. Results from our study reveal a fascination with machine learning and feature engineering models; using machine learning models can lead to superior outcomes. In the medical and diagnostic fields, feature engineering is critical for clinicians making life-changing decisions since it makes it possible to understand the importance and influence of every attribute on tumors subtype categorization It was successful in developing a reliable CAD tool for autonomous lung tissue grouping. Homonymy-based processing of images was more useful for CAD systems than regular texture evaluation.

REFERENCES

[1]M. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: a review," IEEE Rev. Biomed. Eng., vol. 2, 2009.

[2]U. Srinivas, H. S. Mousavi, C. Jeon, V. Monga, A. Hattel, and B. Jayarao, "SHIRC: A simultaneous sparsity model for histopathological image representation and classification," Proc. IEEE Int. Symp. Biomed. Imag., pp. 1118–1121, Apr. 2013.

[3]U. Srinivas, H. S. Mousavi, V. Monga, A. Hattel, and B. Jayarao, "Simultaneous sparsity model for histopathological image representation and classification," IEEE Trans. on Medical Imaging, vol. 33, no. 5, pp. 1163–1179, May 2014.

[4]N. Nayak, H. Chang, A. Borowsky, P. Spellman, and B. Parvin, "Classification of tumor histopathology via sparse feature learning," in Proc. IEEE Int. Symp. Biomed. Imag., 2013, pp. 1348–1351.

[5]H. S. Mousavi, V. Monga, A. U. Rao, and G. Rao, "Automated discrimination of lower and higher grade gliomas based on histopathological image analysis," Journal of Pathology Informatics, 2015.

[6]Jiatai Lin, Guoqiang Han , Xipeng Pan, Zaiyi Liu, Hao Chen , Danyi Li, Xiping Jia , Zhenwei Shi , Zhizhen Wang, Yanfen Cui, Haiming Li, Changhong Liang, Li Liang, Ying Wang, and Chu Han , "PDBL: Improving Histopathological Tissue Classification With Plug-and-Play Pyramidal Deep-Broad Learning"IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 41, NO. 9, SEPTEMBER 2022

[7]N. Coudray et al., "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning", Nat. Med., vol. 24, no. 10, pp. 1559-1567, Sep. 2018.

[8]M. Gehrung, M. Crispin-Ortuzar, A. G. Berman, M. O'Donovan, R. C. Fitzgerald and F. Markowetz, "Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning", Nature Med., vol. 27, no. 5, pp. 833-841, May 2021.

[9]A. Binder et al., "Morphological and molecular breast cancer profiling through explainable machine learning", Nature Mach. Intell., vol. 3, no. 4, pp. 355-366, Apr. 2021.

[10]Y. Fu et al., "Pan-cancer computational histopathology reveals mutations tumor composition and prognosis", Nature Cancer, vol. 1, no. 8, pp. 800-810, Aug. 2020.

[11]M. Y. Lu et al., "AI-based pathology predicts origins for cancers of unknown primary", Nature, vol. 594, no. 7861, pp. 106-110, Jun. 2021.

[12]M. van Rijthoven, M. Balkenhol, K. Siliņa, J. van der Laak and F. Ciompi, "HookNet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images", Med. Image Anal., vol. 68, Feb. 2021.

[13]C. L. Srinidhi, O. Ciga and A. L. Martel, "Deep neural network models for computational histopathology: A survey", Med. Image Anal., vol. 67, Jan. 2021.

[14]M.-J. Tsai and Y.-H. Tao, "Deep learning techniques for the classification of colorectal cancer tissue", Electronics, vol. 10, no. 14, pp. 1662, Jul. 2021.

[15]Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model", Sci. Rep., vol. 7, no. 1, pp. 4172, 2017.

[16]M. Y. Lu et al., "AI-based pathology predicts origins for cancers of unknown primary", Nature, vol. 594, no. 7861, pp. 106-110, Jun. 2021.

[17]L. Cai, J. Gao and D. Zhao, "A review of the application of deep learning in medical image classification and segmentation", Ann. Transl. Med., vol. 8, no. 11, pp. 713, Jun. 2020.

[18]Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," Sci. Rep., vol. 7, no. 1, p. 4172, 2017

[19]P. Sabitha, G Meeragandhi "A Nucleus Based Feature Extraction From Histopathology Images Using CNN For Liver Cancer" 2022 International Conference on Computing, Communication and Power Technology (IC3P)

[20]B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver and J. G. Elmore, "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks", Pattern Recognit., vol. 84, pp. 345-356, Dec. 2018.

[21]Mukhamejan Karatayev, Saltanat Khalyk, Min-Ho Lee,And M. Fatih Demirci "Breast cancer histopathology image classification using CNN"2021 16th International Conference on Electronics Computer and Computation (ICECCO) Year: 2021

[22]M Siva Naga Raju; Dr. B Srinivasa Rao"Classification of Colon Cancer through analysis of histopathology images using Transfer Learning" Published in: 2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)

[23] Mehdi Afshari and H.R. Tizhoosh". A Similarity Measure of Histopathology Images by Deep Embeddings" 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)

[24] Faezehsadat Shahidi, Salwani Mohd Daud,Hafiza Abas,Noor Azurati Ahmad, And Nurazean Maarop"Breast Cancer Classification Using Deep Learning Approaches and Histopathology Image: A Comparison Study"IEEE Access Year: 2020 — Volume: 8 — Journal Article — Publisher: IEEE

[25]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database", Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 248-255, Jun. 2009.

[26]J. Zhang, Y. Xie, Q. Wu and Y. Xia, "Medical image classification using synergic deep learning", Med. Image Anal., vol. 54, pp. 10-19, May 2019.

[27]K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770-778, Jun. 2016.

[28]. C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," Med. Image Anal., vol. 67, Jan. 2021, Art. no. 101813.

[29]S. Wang et al., "RMDL: Recalibrated multi-instance deep learning for whole slide gastric image classification", Med. Image Anal., vol. 58, Dec. 2019.

[30]M. Kowsher,Abdullah As Sami,Nusrat Jahan ,Prottasha,Mohammad Shamsul Arefin;Pranab Kumar Dhar,Takeshi Koshiba"Bangla-BERT: Transformer-Based Efficient Model for Transfer Learning and Language Understanding"IEEE Access Year: 2022, Volume: 10 ,Journal Article,Publisher: IEEE

[31]X. Zhang, X. Zhou, M. Lin and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 6848-6856, Jun. 2018.

[32]Alberto Nogales, Alvaro J. Garcia-Tejedor, Ana M. Maitin, Antonio Perez-Morales, Maria Dolores Del Castillo, And Juan Pablo Romero"BERT Learns From Electroencephalograms About Parkinson's Disease: Transformer-Based Models for Aid Diagnosis"IEEE Access Year: 2022 ,Volume: 10

[33]M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks", Proc. Int. Conf. Mach. Learn., pp. 6105-6114, 2019

[34] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova"BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" Submitted on 11 Oct 2018 (v1), last revised 24 May 2019 (this version, v2)

[35] Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, And Degui Zhi npj"Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction " Digital Medicine volume 4, Article number: 86 (2021)

[36] Jakob Nikolas Kather, Alexander T. Pearson, Niels Halama,Dirk J äger, And Jeremias Krause"Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer" published: 03 June 2019,Nature Medicine volume 25, pages1054–1056 (2019)

[37] Shahid Mehmood, Taher M. Ghazal,Muhammad Adnan Khan, And Muhammad Zubair"Malignancy Detection in Lung and Colon Histopathology Images Using Transfer Learning With Class Selective Image Processing"IEEE Access Year: 2022 — Volume: 10 — Journal Article — Publisher: IEEE

[38]J. N. Kather et al., "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study", PLOS Med., vol. 16, no. 1, Jan. 2019.

[39]A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand and S. M. Mastorides, "Lung and colon cancer histopathological image dataset (LC25000)", arXiv:1912.12142, 2019.

[40]N. Ma, X. Zhang, H.-T. Zheng and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design", Proc. Eur. Conf. Comput. Vis. (ECCV), pp. 116-131, Sep. 2018.

[41]C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, "Inception-v4 Inception-ResNet and the impact of residual connections on learning", Proc. 31st AAAI Conf. Artif. Intell., pp. 1-7, 2017.

[42]I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio, Deep Learning, Cambridge, MA, USA:MIT Press, vol. 1, no. 2, 2016.

[43]C. L. P. Chen and Z. L. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture", IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 1, pp. 10-24, Jan. 2018.

[44]N. Hatami, M. Bilal, and N. Rajpoot, "Deep multi-resolution dictionary learning for histopathology image analysis," Dept. Comput. Sci., Univ. Warwick, Coventry, U.K., Tech. Rep., 2021, doi: 10.48550/arxiv.2104.00669.

[45]S. C. Kosaraju, J. Hao, H. M. Koh, and M. Kang, "Deep-Hipo: Multiscale receptive field deep learning for histopathological image analysis," Methods, vol. 179, pp. 3–13, Jul. 2020.

[46]J. Li et al., "A multi-resolution model for histopathology image classification and localization with multiple instance learning," Comput. Biol. Med., vol. 131, Apr. 2021, Art. no. 104253.