SEGMENTATION-BASED FRACTAL TEXTURE ANALYSIS AND DEEP CONVOLUTION NETWORKS FOR MULTIGRAIN SCORING OF PROSTATE CANCER

Karma M.Fathalla
Computer Engineering, AAST
Alexandria, Egypt
karma.fathalla@aast.edu

Amr Sherif Eltelwany
Computer Engineering, AAST
Alexandria, Egypt
amr.sherif.eltelwany@gmail.com

Amani A. Saad
Prof, Computer Engineering, AAST
Alexandria, Egypt
amani.saad@aast.edu

Abstract

Prostate cancer diagnosis and staging is of paramount importance to effective treatment planning and better prognosis. Computer aided diagnosis and staging can contribute to improving and speeding up these stages, especially with the advances of deep learning. The International Society of Urological Pathology (ISUP) grading of stained Whole Slide histopathological Images (WSIs) can be considered the gold standard for grading. However, WSIs suffer from large size and wide background areas which hinder the learning process. Hence, a segmentation-based fractal analysis approach is applied to address this issue and elect relevant patches to be input to the learning algorithm. EfficientNet Convolution Neural Network (CNN) achieves a promising accuracy of 80.7% and Quadratic Weighted Kappa (QWK) of 95.4%. The proposed approach remedies the size problem of WSIs and improves the grading accuracy using light weight learning models.

***Keywords*: Deep learning; EfficientNet; WSI, segmentation, cancer grading.**

1. Introduction

Computer aided diagnosis (CAD) systems are assistive tools that aid physicians take prompt decisions and plan treatments. CAD is an important field of research that relies on advanced image processing techniques (Halalli & Makandar 2017). Recently, massive strides have been taken towards developing effective CAD systems. The attained progress can be attributed to the adoption of deep learning techniques (Chan et al. 2020). In addition, the integration of manually engineered and automatic deep features achieved considerable success in various diagnostic systems (Liu et al. 2020)(Ibrahim et al. 2020).

A range of image modalities can be used as input to CAD systems. In particular, histopathological image analysis plays a critical role in the diagnosis of many pathologies such as prostate cancer (Komura & Ishikawa 2018). According to the 2020 global cancer statistics (Sung et al. 2021), prostate cancer comprise 14.1% of the new cancer incidence in men, coming second after lung cancer. Also, it presents 6.8% of cancer related deaths in males. Similar findings were reported by the American cancer society(Siegel et al. 2021), where prostate cancer accounts for 26 % of estimated new cases and 11% of estimated deaths. The early detection and precise grading of prostate cancer could help better prognosis and cure(Fong & Djavan 2005).

Artificial Intelligence can help physicians process large volumes of data and discover complex relations rapidly and objectively(Van Booven et al. 2021), hence increase the chances of early detection. For the purpose of early detection and grading prostate cancer, Whole Slide Images (WSI) of prostate histopathology can be used in conjunction with AI methods. However, WSIs vary in size and exhibit very large sizes with billion of pixels, which makes their effective processing through currently available computation models intractable. A common

approach is divide the WSI into patches (tiles) of smaller size to be used for grading. Nevertheless, the choice of the appropriate tiles will dramatically affect the model's performance. Thus, a solution is required to handle this issue.

In this study, we will present a prostate cancer grading model that manages the issues of WSIs. The multigrain grading process relies on Segmentation-based Fractal Textural Analysis (SFTA) and deep Convolution Neural Network (CNN) classification. SFTA is used to exclude image tiles that are of low diagnostic value and hence reduce the image size input to the CNN for final grading.

The remaining of this paper is organized into five sections, where medical background on prostate cancer grading is provided in Section 2. The related work is presented in section 3. In section 4, the proposed approach together with the used dataset are described. The study experiments and results are given in section 5 and conclusions and possible future work are drawn in section 6.

2. Medical Background

Hematoxylin and Eosin H&E stained sections, which are cut from biospsy samples (WSI) are examined for grading and treatment planning. Gleason grading of H&E samples can be considered the gold standard prognostic indicator for prostate cancer (Humphrey 2004). The grading process includes determining and categorizing cancer tissues into Gleason scores of 3,4 or 5. The score is assigned to the tissue sections depending on the growth patterns of the tumor. After the assignment of Gleason scores to the biopsy, an overall ISUP grade (Epstein et al. 2016) is allocated to the sample as shown in Figure 1.

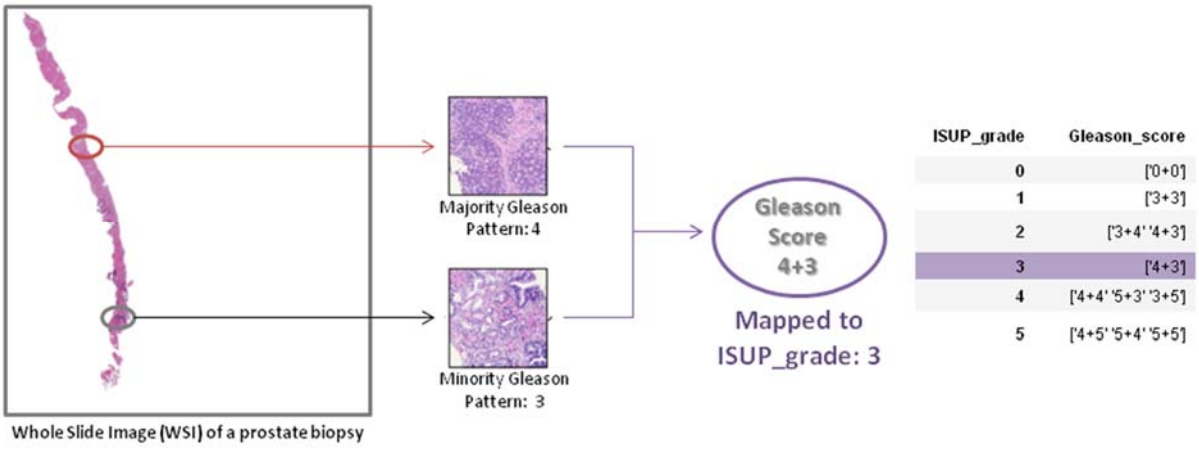


Figure 1 : The ISUP grading process based on Gleason grading of sample patches (tiles) from a WSI prostate cancer biospsy. The example shows that the sample cotain majority patches of Gleason pattern 4 and second most common Gleason pattern 3, presenting a Gleason score of 4+3. The assigned Gleason score is mapped to an ISUP grade of 3 according to International Society of Urological Pathology guidelines provided in table.

Owing to the nature of the grading process, there is a possibility of missing, undergrading or overgrading of cancer tissues. In addition, there is significant inter-observer variability between pathologists. This variability may lead .to giving patients unnecessary treatment or failing to correctly diagnose severe cases. Both cases could result in life detrimental outcomes. Thus, developing a model for prostate cancer assessment based on H&E WSIs of prostate tissue samples, and estimating the severity of the disease became an immensely important task yet a life saver in many cases . For such models to be developed , benchmark datasets are required to allow effective training and results validation.

3. Related Work

Various computer aided diagnosis system are developed for prostate cancer classification and grading. Automated diagnosis systems based on computer vision can help reduce pathologists' inter observer variability and speed up the diagnosis process (Salman et al. 2022). In the wok of Salman et al. (Salman et al. 2022), Yolo object detection algorithm was applied for prostate cancer localization and grade classification on 500 biopsy images split as 450/50 for training and testing respectively. Data Augmentation was used to enlarge the training set from 450 samples to 1776 samples. Two test sets were used, the internal 50 sample dataset and an external 137 samples set. The classification accuracy was 97% for 50 samples and 89% for the external set. Şerbănescu et al. (Şerbănescu et al. 2020) performed gleason score grading on a privately collected dataset of 439 images from 83 patients at the Municipal Clinical Hospital of Cluj-Napoca, Romania. Two deep CNN architectures, namely

AlexNet and GoogleNet, were utilized using transfer learning for grading . The classification task involved distinguishing four cancer grades. Both CNNs attained similar results with mean and standard deviation accuracy of 61.17±7for AlexNet and for GoogleNet results of 60.9±7.4. Other studies adopted more sophisticated approaches as in Singhal et al. study (Singhal et al. 2022). A multi step iterative system is developed where pathologists first annotate the images, then active learning-based data labeling is performed on datasets. Semantic segmentation using Fully Convolutional Network (FCN) is trained for cancer grade determination. After which, the system segmented glands based on Gleason patterns 3, 4, and 5. Following that step, the unlabeled images are input to the trained FCN to give a degree of uncertainty to each image. Then, a pathologist label the most uncertain samples and the samples are added to the training set. This process is iteratively repeated until no samples have an uncertainty above a specific threshold. The model constructed is used for prediction on the test set. The system achieved an accuracy of 89.4% on the internal test set. In Yang et al. (Yang & Xiao 2021), a subset of the tiles are selected based on mean pixel value. A Multi-Channel and Multi-Spatial (MCMS) attention mechanism is proposed to be added to any CNN architecture to improve feature extraction and aid the backbone CNN to focus on more relevant areas of the image. ResNet-50 and ResNet -101 are used as the backbone CNNs and compared to the performance of the baseline Convolutional Block Attention Module (CBAM). The MCMS-based model scored the highest results with and accuracy of 71.8% on PANDA dataset. The adopted approaches either use the WSI as input and rely on heavy Deep CNN architecture for feature extraction, or perform grid or segmentation based sampling to choose the relevant tiles. However, the available grid sampling approach rely on primitive aspects such as mean pixel value, which is not effective. On the other hand, the segmentation procedure might lead to loss of details in the surrounding tissue of the Region of Interest. Hence, in this study an effective approach for tile selection is proposed coupled with a light weight CNN architecture for grading.

4. Materials and Methods

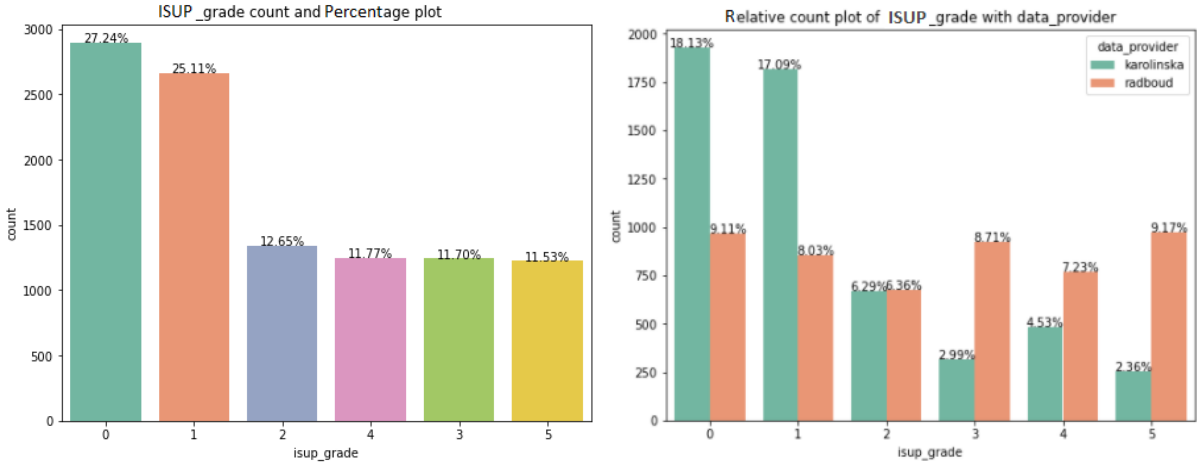
A brief description of the dataset is given, followed by the findings from data exploration. Then, the stages for prostate cancer ISUP grading are described in detail.

4.1. *Dataset Description, Exploration and Wrangling*

In this study, Prostate cANcer graDe Assessment (PANDA) dataset (Bulten et al. 2022) is used for model training and validation. PANDA dataset is considered the largest WSI prostate cancer dataset publicly available for experimentation. . It is provided by two institutes namely Radbound University (Anon n.d.) and Karolinska Institute (Anon n.d.). The Dataset consists of 10616 whole slide image, each image has three size levels , different dimensions across the same level. There are 10516 masks available. Each image has an ISUP grade where the distinct number of ISUP grades is 6 (0-5), and each image has a Gleason score. Each individual image on average is about 25,000pixel x 25,000 pixel RGB. In our experiments, we use a medium resolution input images (4x smaller than the highest resolution).

Dataset Exploration:

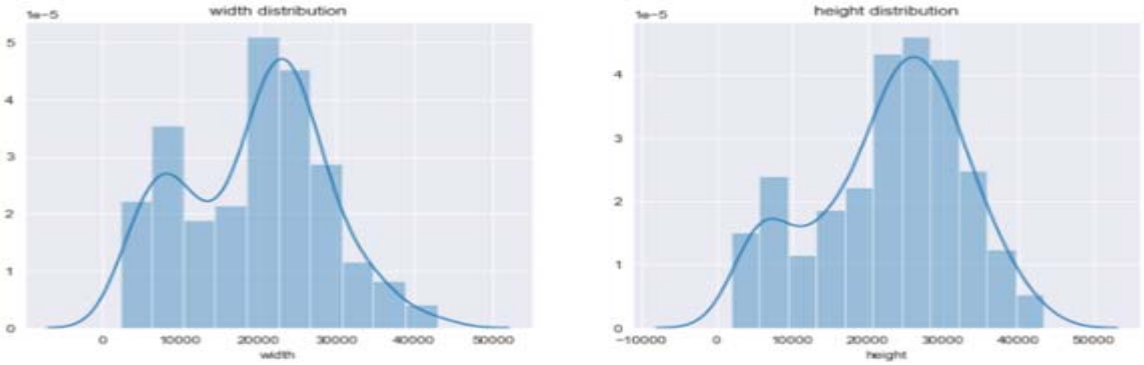
Karolinska Institute contributed to the dataset with 51.39 %and the remaining 48.61% were provided by. Radbound University. First, the ISUP grade distribution is explored to investigate the need for imbalance class handling. The ISUP grade (class) distribution show skewness as ISUP grades constitute around 52 % of the overall sample, and the remaining 48% is almost uniformly distributed across the remaining grades (2-5) presenting about 12% each. Despite the skewed distribution, it is not considered as highly imbalanced according to the study of Hassanin et al. (Hasanin et al. 2019), which considers a minimum ratio of 100:1 to be highly imbalanced. Hence, the need for imbalance handling is eliminated.



(a) **(b)**

Figure 2: ISUP grade distribution (a) across the complete dataset (b) segmented per data provider

Another aspect that is explored is the dimensionality of the WSIs. Figure 3 shows the histogram of the widths and heights of the avaialble WSIs. The counts show that the mode values lie in the range 20,000 to 30,000 pixel, which emphasizes the dimensionality issue. Another illustration is shown in Figure 4 , where the sizes of the WSIs can be divided into three levels of 0, 1 and 2. Each image is available in the three sizes with level 0 denoting the largest size level. The depicted sizes clarify the dimensionality isuue of WSI, where the smallest level is composed of thousands of pixels. Hence, a solution is required that maintains informative patches (tiles) and discards tiles of low information content.



(a) **(b)**

Figure 3: Histogram of the (a) widths and (b) heights of the WSIs.

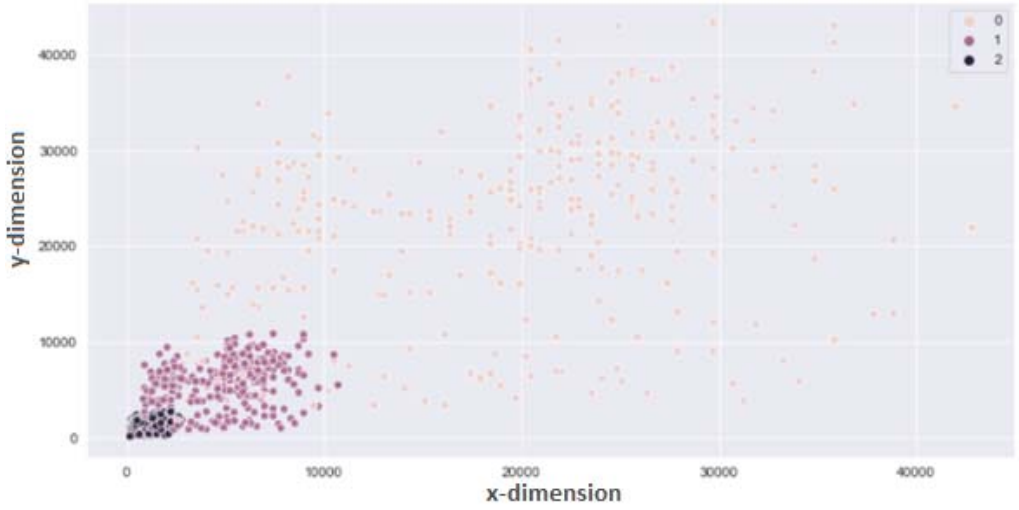


Figure 4: The dispersion of the WSI across the x - and y- dimension divided into three size levels 0, 1 and 2

Data Wrangling

First, mislabeled images are checked and removed from the dataset. Mislabeled images are identified as images with wrong mapping of gleason score and ISUP grade. It was found that only one image was wrongly mapped , a gleason score '4+3' was wrongly mapped to ISUP grade 2. Also, images that contain only background (white images) are eliminated using color histogram on the pixels and removing the images that contain only one value in the histogram. As a result of this procedure, four images are cleaned from the dataset. In addition, Removing images with no or suscpious masks that have high probability to be mislabbeled and Images without masks are removed from the dataset. In total, 100 images were found without masks and 85 image seemed suscpious as they are labeled as cancerous and no cancerous pixels can be found on the masks

Finally, during data wrangling some images are found having pen marks as can be seen in Figure 5. These Images are removed as they can be considered as noise to the traning algorithm, 475 images are removed from the dataset by using a simple histogram comparison for the values of pixels and removing images with outlier values in its histogram. Manual inspection is performed before complete removal.

Figure 5: Examples of pen marked images that were removed during data wrangling

4.2. *Proposed Model for Prostate Cancer Grading*

In this section, the main phases of the proposed model are described. Figure 6 shows the flow of the prostate cancer grading process. The WSIs are partitioned into tiles, each of size n x n. Every tile is processed using Segmentation-based Fractal Textural Analysis (SFTA) and assigned a rank accordingly. The top ranked tiles are input to the Deep CNN learning architectures, which produces a prostate cancer grade. The use of SFTA is described in further details below. In addition, a brief description of the employed CNN architectures is provided.

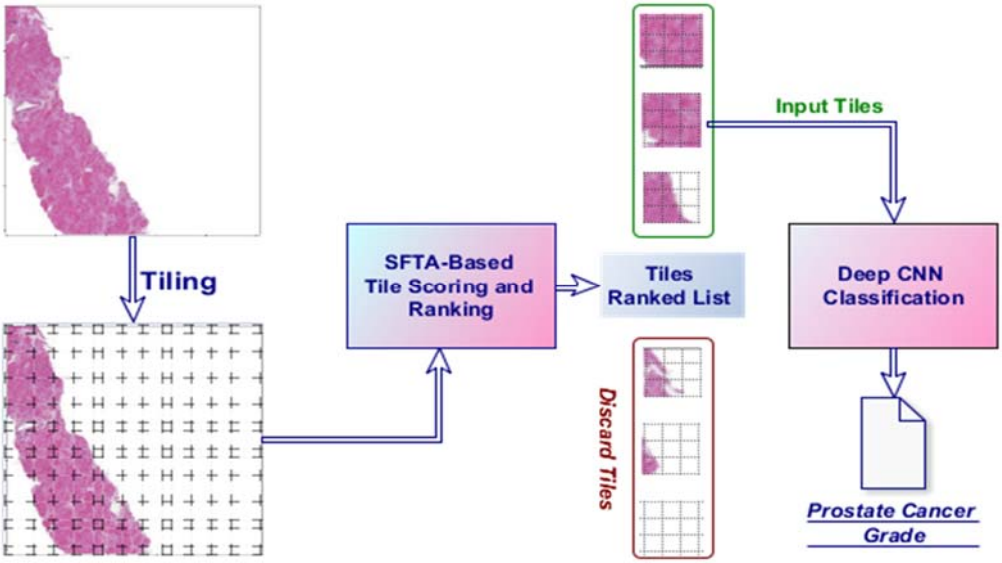


Figure 6: Prostate Cancer Grading process flow using Segmentation-based Fractal Textural Analysis (SFTA) and Deep CNN Classification

SFTA-based Tile Scoring and Ranking

SFTA is a feature extraction approach based on textural descriptors. Textural features of an image can be considered of the most discriminative descriptors, which can be used to differentiate between prostate cancer grades. SFTA is used for feature extraction to characterize WSI grades as it is known for its robustness and low computation cost (Costa et al. 2012). The SFTA features extraction method comprises two main steps, which are:

- Multiple Binary images formulation.
- Feature vector generation from the boundary images

For binary images formulation, Two-Threshold Binary Deterioration (TTBD) is applied on input grayscale image I(i,j) for binary decomposition. A set of thresholds are calculated based on multilevel Otsu algorithm (Otsu 1979), which aims at finding a threshold that minimizes the input image intra-class variance. The number of thresholds to be computed is input as a user defined parameter (n_t). Otsu algorithm is applied recursively until reaching the number of thresholds required. Then, a series of binary images are produced using pairs of the computed thresholds as lower bounds (t_{lb}) and upper bounds (t_{ub}) thresholds according to equation (1).

$$I_b(i,j) = \begin{cases} 1 & if\ t_{lb} < I(i,j) < t_{ub} \\ 0 & otherwise \end{cases} \quad (1)$$

where in this study, I(i,j) represent each tile in the WSI.

After creating the set of binary images, a feature vector is generated as the binary images' size (pixel count), mean gray level and boundaries' fractal dimension. The fractal dimension determines the complexity of the boundary images, Δ(i,j) . The boundary image is generated using the following equation (2):

$$\Delta(i,j) = \begin{cases} 1 \text{ if } \exists(i',j') \in N_8[(i,j)]: \\ \quad I_b(i',j') = 0 \wedge \\ \quad I_b(i,j) = 1, \\ 0, \text{otherwise} \end{cases} \quad (2)$$

where $N_8[(i,j)]$ is 8 neighborhood pixels connected to pixel (i,j). Δ(i,j) is assigned 1 if the pixel at position (i; j) in the respective binary image I_b(i; j) has the value 1 and having at least one neighboring pixel with value 0. Otherwise, Δ(i,j) takes the value 0. The fractal dimension fd_i is computed from each border image using the box counting algorithm (Costa et al. 2012).

In our study, the factal dimensions of each tile are used to compute a rank score for each tile. The rank score is computed as ratio between average fractal dimensions of each tile to the overall mean fractal dimension of the whole WSI. The scores are sorted and the top *m* are selected to be input the Deep CNN architectures for classification.

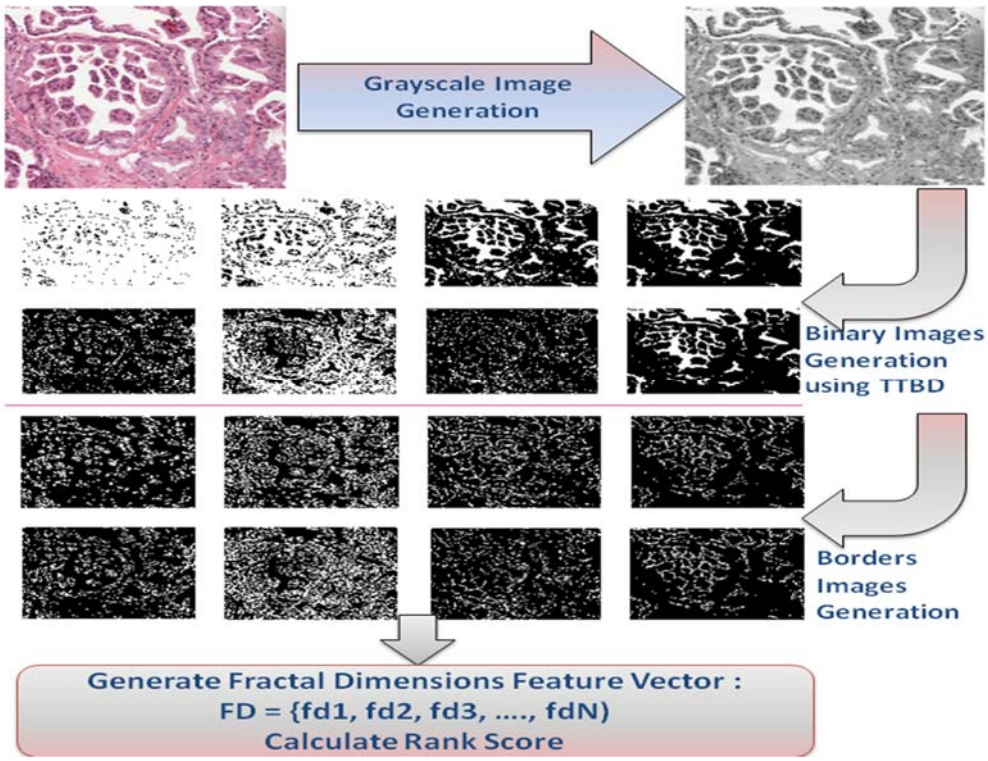
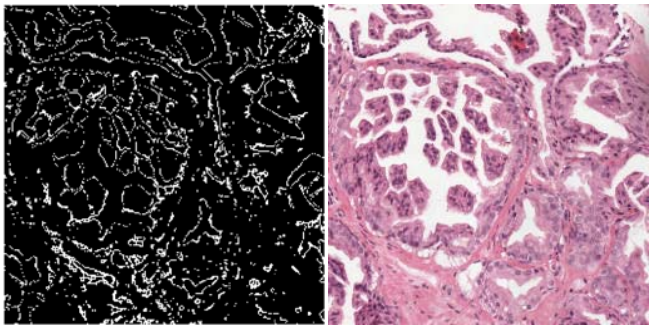
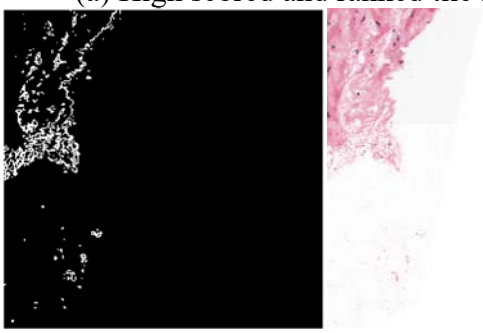


Figure 7: SFTA feature extraction process and rank score calculation from fractal dimension features



(a) High scored and ranked tile sample



(b) Low scored and ranked tile sample

Figure 8 SFTA output (border fractal dimension) for two sample tiles (a) and (b) from a single patient WSI

Deep CNN Classification

The highly ranked tiles are used to construct a single image and input to Deep Convolutional Neural Networks (CNN) for cancer grading. CNN is a neural network with multiple layers architecture (Dargan et al. 2020). CNN can extract automatic discriminative features which have some invariance properties (e.g. translation invariance) (Krizhevsky et al. 2017). It consists of three main layers which are convolution layers, pooling layers and fully connected layers. The early convolution layers of the architecture are used for extracting local low-level features from the raw input while the deeper convolution layers of CNN are used for combining features together to generate global high-level features (Dargan et al. 2020). The pooling layers are used to down sample the dimensionality of the extracted feature. The fully connected layers form an ANN network where each neuron in the previous layer is connected to all the neurons in the current layer. The total number of fully connected neurons in the final layer determines the number of classes.

The advantages of CNNs include that they are well suited for end-to-end learning that generates automatic features from the raw data without any a priori feature selection. Moreover, CNNs scale well to large datasets. The disadvantages of CNNs include that they may output false predictions with high confidence, may require a large amount of training data, may take longer to train than simpler models, and involve a large number of hyper parameters such as the number of layers or the type of activation functions, the limited interpretability of the constructed models. Some of famous CNN are as follow:

- **EffcientNet model:** A compound scaling method was proposed for better performance (accuracy) and efficiency (model parameters). The scaling method used a fixed set of coefficients to scale width, depth and resolution. The baseline network EfficientNet B0 was produced and further scalings of it were constructed as EfficientNets B1-B7 (Tan & Le 2019). The total number of layers in EfficientNet-B0 is 237 and in EfficientNet-B7 the total is 813.
- **Visual Graphic Group Net (VGG Net) model:** This net was developed by the technicians at the Visual Graphics Group from the Oxford and is in pyramid shape. The model consists of the bottom layers which are wide and the top layers are deep. There are 2 versions of VGG which are VGG-16 and VGG-19. VGG-16 is a combination 13 convolutional layers and three fully connected layers (Simonyan & Zisserman 2014).
- **Xception model :** It has 71 layers. It started by two convolution layers followed by depth separable convolution layers, four convolution layers and dense layer (Chollet 2017).
- **NASNet Model :** Neural Architecture Search (NAS) for cells Network comprises a set of convolutional cells (blocks), namely normal cells and reduction cells. The overall architecture is searched by reinforcement learning (Zoph et al. 2017).

5. Results and Discussion

SFTA-based tile selection was implemented using MATLAB R2019a. CNN models implementation, training and results were done using python language with Numpy, Tensorflow2.0, Keras and OpenCV packages (Anaconda Python). Experiments were conducted on Google Colab. Google Colab consists of 1xNvidia Tesla K80, 12.6 GB RAM and 320 GB Disk space. All Deep Learning architectures were trained for 30 epochs from scratch using Adam optimizer with starting learning rate of 0.001.

5.1. *Performance Metrics*

The performance of the proposed system is evaluated using a set of established metrics. Different measures are used for each stage.

For the SFTA-based Tile Scoring and Ranking stage, ANalysis Of VAriance test (Dawson & Trapp 2004) is performed to validate the presence of a significant difference between the ranking scores of selected and unselected tiles. One-way ANOVA compares two or more group means and the grand mean (selection score) to determine whether the tested groups are drawn from the same population. If the ANOVA result is significant the null hypothesis (all samples are drawn from same population) is rejected. Three assumptions are made when applying one-way ANOVA which are: the values of the dependent variable (selection score) is normally distributed, the population variance is the same in all groups and the recorded observations are random and independent. We will present three outputs of ANOVA namely the p-value, F criteria and sample box plots in case of significant findings. In the conducted experiments, a p-value <0.01 was considered as significant.

For the final grading, Accuracy (Acc) and Quadratic Weighted Kappa (QWK) are used . ACC measures the percentage of correct predictions over the total number of tested samples , as shown in Equation (3). The number of correct predictions is the summation of True Positive (TP) and True Negative (TN) divided by the overall test set size TS_t.

$$Acc = \frac{TP + TN}{TS_t} \tag{3}$$

On the other hand, QWK is used to measure the degree of agreement between the made predictions and the ground truth. QWK calculation relies on O, W and E matrices of size NxN for a N multiclass classification problem. Element O_{ij} in O confusion matrix represents number of ISUP grade i that receives a predicted value j. The E matrix corresponds to the expected outcomes, calculated as the outer product between the actual rating's histogram vector and the predicted rating's histogram vector. W matrix assigns a weight w_{ij} for every prediction - actual outcome pair depending on the difference between them. Equations (4) and (5) show the calculation of W and QWK..

$$w_{ij} = \frac{(i-j)^2}{(N-1)^2} \tag{4}$$

$$QWK = 1 - \frac{\sum_{ij} w_{ij} O_{ij}}{\sum_{ij} w_{ij} E_{ij}} \tag{5}$$

5.2. *Performance Evaluation*

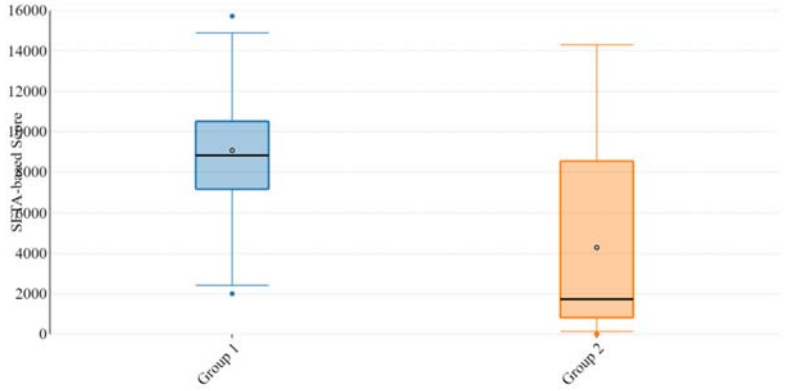
ANOVA Analysis

Samples of WSIs are randomly chosen to apply ANOVA test. Each WSI is divided into (tiles) patches of size 256 x 256. The highest ranked tiles (top 10) and the least 10 scored tiles for each WSI are input to the ANOVA test. The output of the ANOVA test are presented in Table 1. Both p-value and F-criteria show significant difference between the two test groups, which leads to the rejection of the null hypothesis.

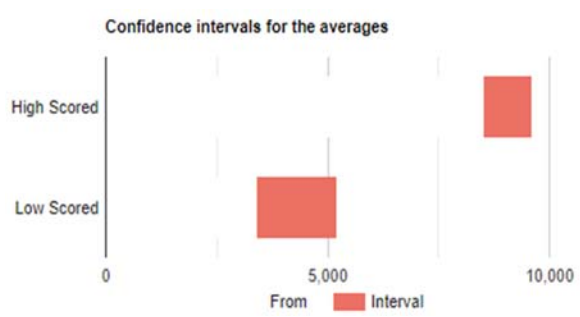
ANOVA Test Outputs	**p-value**	**F-criteria**
	2.98E-05	4.67
Null Hypothesis	Rejected	Rejected

Table 1: ANOVA test output comparing selection ranking scores of the top 10 tiles and the least scored 10 tiles of WSIs.

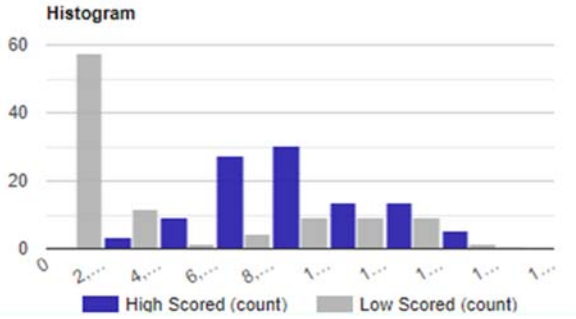
In addition, Figure 9 shows the box plots, mean confidence interval and histogram of binned ranking scores. All the illustrated plots conform the p-value and F-criteria conclusion of rejecting the null hypothesis. Such finding outlines the discriminating ability of SFTA-based ranking score to create distinguishable relevant and irrelevant tiles, which is further clarified by the grade scoring classification results.



(a) ANOVA analysis box plot for Group 1 (high ranked) and Group 2 (low ranked)



(b) Averages Confidence Interval



(c) Histogram of binned scores

Figure 9 Illustration of the significant differences in terms of variance, average and histogram counts between sample high ranked Vs low ranked tiles high scored

Grade Scoring Classification Results

In the first experiment, scaled full WSI are input to the described CNN architectures. THE WSI is scaled to 1024 x 1024. This experiment aims to highlight the important role of tile selection and its effect on the classification results. Figure 10. depicts low performance of the CNN architectures, due to the presence of irrelevant background areas in the WSI and loss of details due to scaling. Notably, the highest accuracy of around 62% is attained by EfficientNet - B0. Therefore, EfficientNet is used for the following experiments.

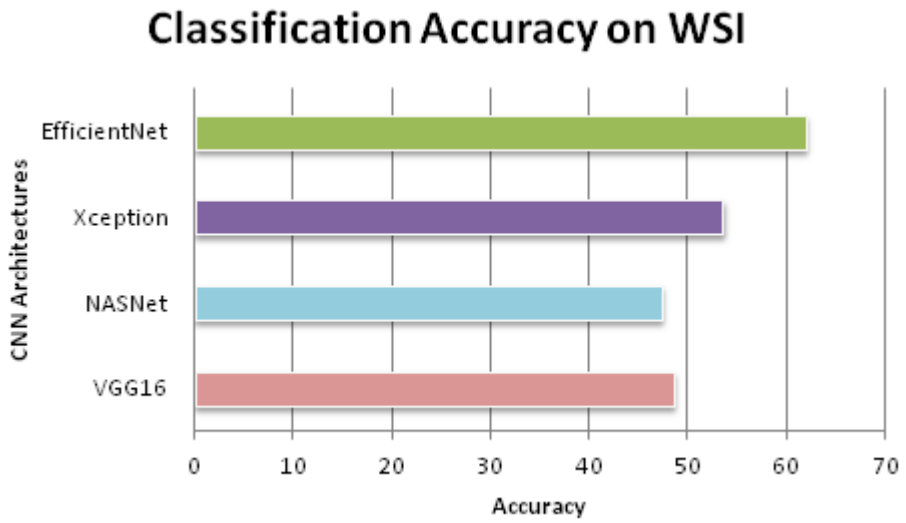


Figure 10: CNN classifications accuracy comparison on scaled full WSIs.

The performance of EfficientNet scaling variants, namely B0, B1, B2, B3, B4 and B5 is tested coupled with a subset of the selected tiles. The highest ranked 36 tiles are used for the following experiment, the results are reported in Figure 11.

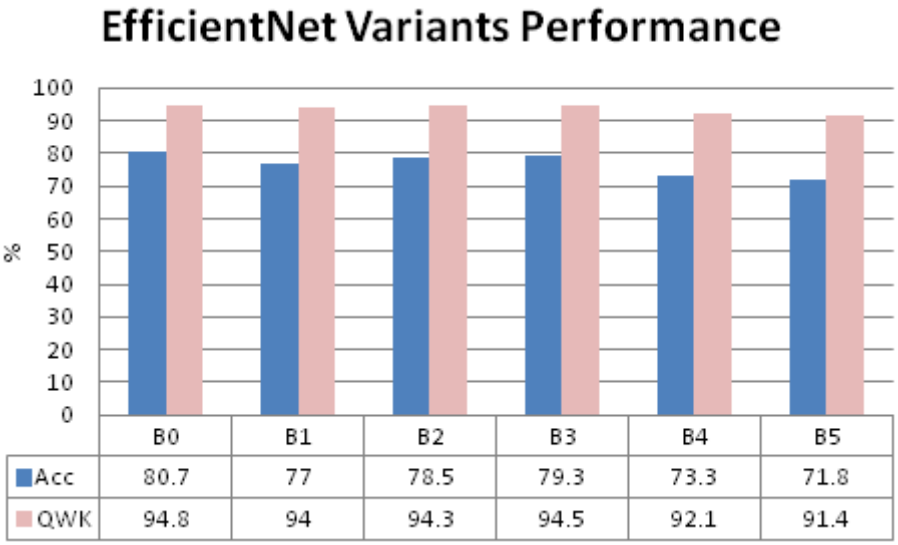


Figure 11: Classification performance comparison between EfficientNet variants using 36 tiles as input to the CNN architectures.

An interesting finding is the performance of Efficient B0 surpasses the performance of B5 in terms of Acc and QWK with a difference of 8.9% and 3.4% respectively. This might be attributed to the vanishing gradient problem. The vanishing gradient problem occurs to mapping of large input space to smaller one (Heng & McColl 2021), which is further exacerbated by the scaling of EfficientNet architectures variants (Ali et al. 2022) when comparing B5 and B0.

EffiicientNet B0 is used in the remaining experiment, as it presents the highest performance and the lightest weight architecture. Another aspect that is studied is whether the number of selected tiles impacts the output ACC and QWK. Figure 12 depicts the effect of changing the number of selected tiles on the performance of EffiicientNet B0. The tested grids are 6x6, 7x7 and 8x8. The results show minimal variation between the different grid sizes with a maximum difference of 1% Acc and 0.6% QWK. The obtained results show comparable performance. However, the training time for 36 tiles is lower than its counterparts.

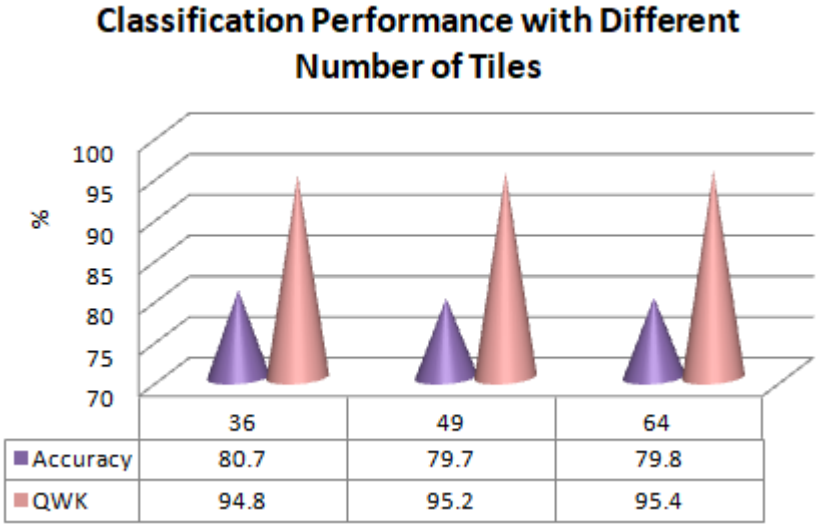


Figure 12: Classification performance comparison of EfficientNet-0 using different number of input tiles.

The multigrain approach achieves better results than the recent work of Yang et al.(Yang & Xiao 2021), which scored an Acc of 71.7% and QWK of 85.04%. It is worth noting that Yang et al. adopted a complex dual attention mechanism to assist CNN learning, whereas a light weight CNN namely EfficientNet B0 is employed in this study approach.

6. Conclusion

In this study, a multigrain scoring approach is presented for prostate cancer ISUP grading. The presented approach employs SFTA to score and rank WSI tiles (patches) to eliminate irrelevant and low-value tiles. This step reduces

computation time and improves the classification (grading) accuracy. A range of CNN architectures are tested, EfficientNet B0 attains the highest accuracy of 80.7 % with 36 of the top ranked tiles and the highest QWK of 95.4% with 64 tiles. The results show promising results that surpassed the state of the art in the literature. Refinement to the tile selection approach can be attempted through the inclusion of medical expertise to improve the accuracy. In addition, different CNN architectures can be combined with the presented approach.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ali, K. et al., 2022. Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer. *Neuroscience Informatics*, 2(4), p.100034. Available at: https://www.sciencedirect.com/science/article/pii/S2772528621000340.
[2] Anon, Karolinska Institutet : Department of Medical Epidemiology and Biostatistics. Available at: https://ki.se/en/meb [Accessed July 13, 2022a].
[3] Anon, Radboud university medical center. Available at: https://www.radboudumc.nl/en/research [Accessed July 13, 2022b].
[4] Van Booven, D.J. et al., 2021. A Systematic Review of Artificial Intelligence in Prostate Cancer. *Research and reports in urology*, 13, pp.31–39.
[5] Bulten, W. et al., 2022. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nature Medicine*, 28(1), pp.154–163. Available at: https://doi.org/10.1038/s41591-021-01620-2.
[6] Chan, H.-P., Hadjiiski, L.M. & Samala, R.K., 2020. Computer-aided diagnosis in the era of deep learning. *Medical Physics*, 47(5), pp.e218–e227. Available at: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13764.
[7] Chollet, F., 2017. Xception: Deep Learning With Depthwise Separable Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[8] Costa, A.F., Humpire-Mamani, G. & Traina, A.J.M., 2012. An Efficient Algorithm for Fractal Analysis of Textures. In *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*. pp. 39–46.
[9] Dargan, S. et al., 2020. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Archives of Computational Methods in Engineering*, 27(4), pp.1071–1092. Available at: https://doi.org/10.1007/s11831-019-09344-w.
[10] Dawson, B. & Trapp, R.G., 2004. *Basic & Clinical Biostatistics 4/E (EBOOK)*, McGraw-Hill Education. Available at: https://books.google.ne/books?id=p6hu-qU2zpsC.
[11] Epstein, J.I. et al., 2016. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *The American journal of surgical pathology*, 40(2), pp.244–252.
[12] Fong, Y.K. & Djavan, B., 2005. Early detection of prostate cancer. *Reviews in urology*, 7(1), pp.63–64.
[13] Halalli, B. & Makandar, A., 2017. Computer Aided Diagnosis - Medical Image Analysis Techniques. In C. M. Kuzmiak, ed. *Breast Imaging*. Rijeka: IntechOpen. Available at: https://doi.org/10.5772/intechopen.69792.
[14] Hasanin, T. et al., 2019. Severely imbalanced Big Data challenges: investigating data sampling approaches. *Journal of Big Data*, 6(1), p.107. Available at: https://doi.org/10.1186/s40537-019-0274-4.
[15] Heng, L. & McColl, B. eds., 2021. *Mathematics for Future Computing and Communications*, Cambridge University Press.
[16] Humphrey, P.A., 2004. Gleason grading and prognostic factors in carcinoma of the prostate. *Modern Pathology*, 17(3), pp.292–306. Available at: https://doi.org/10.1038/modpathol.3800054.
[17] Ibrahim, M.R., Fathalla, K.M. & M. Youssef, S., 2020. HyCAD-OCT: A Hybrid Computer-Aided Diagnosis of Retinopathy by Optical Coherence Tomography Integrating Machine Learning and Feature Maps Localization. *Applied Sciences*, 10(14). Available at: https://www.mdpi.com/2076-3417/10/14/4716.
[18] Komura, D. & Ishikawa, S., 2018. Machine Learning Methods for Histopathological Image Analysis. *Computational and Structural Biotechnology Journal*, 16, pp.34–42. Available at: https://www.sciencedirect.com/science/article/pii/S2001037017300867.
[19] Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM*, 60(6), pp.84–90. Available at: https://doi.org/10.1145/3065386.
[20] Liu, Y. et al., 2020. Using CNN With Handcrafted Features for Prostate Cancer Classification. In *2020 5th International Conference on Automation, Control and Robotics Engineering (CACRE)*. pp. 636–641.
[21] Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp.62–66.
[22] Salman, M.E. et al., 2022. Automated prostate cancer grading and diagnosis system using deep learning-based Yolo object detection algorithm. *Expert Systems with Applications*, 201, p.117148. Available at: https://www.sciencedirect.com/science/article/pii/S0957417422005425.
[23] Şerbănescu, M.S. et al., 2020. Automated Gleason grading of prostate cancer using transfer learning from general-purpose deep-learning networks. *Romanian journal of morphology and embryology = Revue roumaine de morphologie et embryologie*, 61(1), pp.149–155.
[24] Siegel, R.L. et al., 2021. Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians*, 71(1), pp.7–33. Available at: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21654.
[25] Simonyan, K. & Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. Available at: https://arxiv.org/abs/1409.1556.
[26] Singhal, N. et al., 2022. A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. *Scientific Reports*, 12(1), p.3383. Available at: https://doi.org/10.1038/s41598-022-07217-0.
[27] Sung, H. et al., 2021. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), pp.209–249. Available at: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21660.
[28] Tan, M. & Le, Q. V, 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Available at: https://arxiv.org/abs/1905.11946.
[29] Yang, B. & Xiao, Z., 2021. A Multi-Channel and Multi-Spatial Attention Convolutional Neural Network for Prostate Cancer ISUP Grading. *Applied Sciences*, 11(10). Available at: https://www.mdpi.com/2076-3417/11/10/4321.
[30] Zoph, B. et al., 2017. Learning Transferable Architectures for Scalable Image Recognition. Available at: https://arxiv.org/abs/1707.07012.

Authors Profile

Karma M. Fathalla, attained her Master of Science degree from Arab Academy for Science and Technology (AAST), Egypt, while she has received her Ph.D. in machine learning from Aston University, UK. Her research work included data mining, signal and image processing, pattern recognition. During her studies as a researcher, she presented her work in several reputable conferences and high impact factor journals. Currently, her research interests are multimodal signal processing , health informatics and machine learning. Karma M. Fathalla is holding an assistant professor position at the department of Computer Engineering, AAST, Egypt.

Amr Sherif Eltelwany, achieved his BSc. in Computer and Communications Engineering from Alexandria University, Egypt. He graduated third on his 2016 class. He is currently a Master of Science student in the Computer Engineering department at Arab Academy for Science and Technology. He is a certified data engineer with more than 6 years experience on data product development, which includes around four years of experience in big data analytics

Amani Anwar Saad is currently a Professor of Computer Engineering in the Arab Academy for Science and Technology in Egypt. She received her BSc in 1988 and her MSc in 1991 from the Department of Computer Science and Automatic Control in the Faculty of Engineering Alexandria University in Egypt. She received her PhD in Computer Engineering from Milan University Italy in collaboration with Alexandria University in 1996. Prof Amani Saad has been working in Alexandria University as Teaching assistant, Assistant Professor and Associate Professor from 1989 till 2006. She was the Head of the Computer Engineering Department at Pharos University in Alexandria since it was established in 2006 till 2010. In October 2010 she joined AAST as a Professor of Computer Engineering. Professor Amani Saad has published more than 50 technical papers and supervised more than 50 graduation projects. Her current research interests include, Machine Intelligence and medical applications, Spatial and Temporal Databases, Context Aware systems and Software Engineering.