

Exploration of Heterogeneous Activation Functions on Oral histology Images through Convolutional Neural Network Approach

Jenifer Blessy. J¹, Sornam.M², Prabhakaran Mathialagan³

¹Department of Computer Science University of Madras Chennai – 600025, India

²Department of Computer Science University of Madras Chennai – 600025, India

³School of CS & IT, Department of Computer Application, Jain (Deemed-To- Be) University, Bangalore –560069, India

Cite this paper as: Jenifer Blessy. J, Sornam.M, Prabhakaran Mathialagan (2024) Exploration of Heterogeneous Activation Functions on Oral histology Images through Convolutional Neural Network Approach. *Frontiers in Health Informatics*, 13 (7), 17-26

Abstract:

In dentistry, Oral Squamous Cell Carcinoma (OSCC) is the most serious and potentially fatal condition. The patient's probability of survival would rise with early diagnosis and medication. Compared to other digital imaging types, histopathological specimens are more helpful in determining the presence of oral malignant images. Haematoxylin and eosin are two stains that can be used to differentiate between cancerous and normal cells. Malignant cells can be distinguished by their appearance, size, spread, and morphological traits. The factors that need to be considered are marginal adhesion, homogeneity of cell shape, and lumpy size. Given the complex and unpredictable nature of the images, OSCC classifying raises serious challenges. A robust strategy is needed to handle these problems and ensure reliable classification. To generate automated systems, the OSCC histopathology images are processed efficiently. To discern between images that are malignant versus those that are not in the histopathology image dataset, a novel Convolutional Neural Network (CNN) model with distinct activation functions is developed. Activation functions are crucial in cancer and non-cancer scenarios when classifying images. The flattened layer executes binary image sorting employing the Sigmoid activation function. The proposed CNN model is enhanced using ReLU, LeakyReLU, Tanh, Swish, PReLU, GeLU, and ELU in three hidden layers of CNN. The swish activation function offers superior categorization with low loss and increased accuracy. The results highlight activation function selection's significance in improving CNN diagnostic accuracy for OSCC and providing vital novel data to develop reliable and effective medical image classification systems.

Keywords: Oral Squamous Cell Carcinoma, Histopathological images, CNN, activation functions, accuracy, loss.

1. Introduction

The sixth most prevalent form of cancer globally is oral cancer. Oral cancer is sometimes misunderstood as an oral ulcer or gum issue. So, the identification of oral tumor is a bit challenging. Tobacco, excessive alcohol, smoking, weaker immune system, prolonged exposure to the sun, and Human papillomavirus are some of the reasons for oral cancer. Oral cancer not only occurs in the mouth but also the head and neck region. The oral cavity's cells remain continuously altered. The variety of cells in which the cancer develops determines the type of mouth cancer. While considering oral cancer, Tumour in the Oral Squamous Cell, Oral Verrucous Carcinoma, and Oral melanoma are the three major types of Oral cancer. The overwhelming majority of oral cavity malignant tumors are squamous cell carcinomas. These cells are composed of thin, flat cells that line the lips and the cavity of the mouth; however, mouth cancers naturally develop. Squamous cell carcinoma comprises a vast majority of oral benign tumors. Verrucous is the rarest type of oral cancer. Melanoma cancer occurs anywhere on the skin, nose, and mouth. Oral Cancer may arise in the

upper palate, hard palate, tongue, gum, floor of the mouth, and lip. If unnoticed for a long period, the cancer cells may navigate to the different parts of the body.

Different types of Oral cancer imaging include Histopathologic, clinical imaging, Narrow band imaging, MRI, PET, CT, CBCT, X-ray, Optical Coherence Tomography (OCT), Confocal Microscopy, and Intraoral Imaging. This work concentrates on microscopic or histopathologic images. Beyond various types of imaging, histopathologic images are more helpful in identifying oral cancerous images. Histopathology images are used for cancer recognition and diagnosis. Hematoxylin (Basic) and Eosin (Acidic) dyes are the two predominant stains that help to identify cancerous cells from normal ones. The hematoxylin-eosin staining technique is used, particularly in cancer diagnosis, to differentiate between different tissue types and their morphological modifications. Haematoxylin, a rich blue-purple stain, and Eosin, a pink stain, highlight the nucleic acid sequences. A normal tissue has variable degrees of pinkish staining in the cytoplasm and extracellular matrix, while the nuclei are stained blue. The cell size, spread, appearance, and morphological features such as border size, density, and shape can substantiate the cancerous cells and nuclei. This leads to the identification of the epithelial layer's cells from other tissues. The tissues for the images are obtained by a process called biopsy. Figure 1 and 2 [1] depicts the view of biopsy cell images under the microscope.

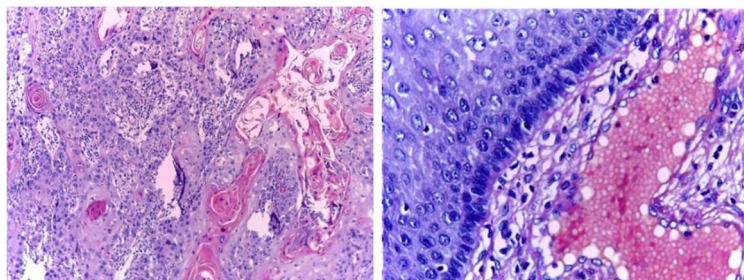


Fig: 1 Cancerous Histopathologic Image Fig: 2 Normal Histopathologic Image

Detection and Identification are the two major phases in reducing the mortality rate in the case of oral cancer. Early and fast detection with more accuracy is one of the major tasks. To detect oral cancer, deep learning, and machine learning models have been developed. The three primary layers of an artificial neural network (ANN) are the input layer, the hidden layer, and the output layer. Using random weights and activation functions, hidden layers are primarily responsible for feature extraction and classification between these layers. ANN works with n number of neurons. The weights can be triggered and activated automatically or manually. Activation functions in the neurons are completely responsible for the output of the desired model. Activation functions can both affect the network and make it effective. Selecting the suitable activation for every individual model is more important. The computerized analysis minimizes time and reduces the need for human experience in identifying malignant and non-malignant cells. To classify oral squamous cell carcinoma using histopathologic images, it is vital to find the most appropriate activation functions, which would impact the framework of the proposed CNN model. Different activation functions have different functionalities.

2. Literature Review

In medical imaging research, CNNs and other deep learning techniques are currently state-of-the-art. Loris Nanni et al. [2], the study set out to evaluate a few of the latest deep-learning methods using data and images from the medical field. To determine how to accomplish this, they analyzed the results of CNN ensembles constructed by replacing activations in the ReLU layers with activations from a variety of activation functions, including six new activation functions that are introduced here as a first (2D Mexican ReLU, TanELU, MeLU + GaLU, Symmetric MeLU, Symmetric GaLU, and Flexible MeLU). Tests were performed using two different networks, VGG16 and ResNet50, on fifteen difficult image data sets representing a range of activities. Experiments also show that the individual architectures are not superior to one another based on histopathologic images.

Kit Wong et al [3], investigate the effects of weight initialization in an artificial neural network. Five various activation

functions are paired with the Nguyen-Widrow weight initialization, random initialization, and Xavier initialization method. The input layer, hidden layer, and output layer of a feedforward neural network are the subjects of this research. Based on their highest obtained loss rate in training, the matched combination of activation functions and weight initialization methods is analyzed, tested, and compared. Several integer variables in the Cancer dataset. The results reveal that the Xavier weight initialization with an activation function that is linear works best, with 0.04909 as the lowest loss score. On average, the Nguyen-Widrow and Xavier weight initializations perform superior on the dataset compared to the random weight initialization.

K Vijayakumar et al [4], for the classification of breast cancer, a completely novel deep Feedforward NN model with four Activation Functions (AF) has been proposed: It has four hidden layers such as Swish, Leaky ReLU, ReLU, and Sigmoid. The study has two goals. First, this work contributes to a deeper comprehension of DNN with varying AFs across layers. Secondly, studies are being conducted to investigate more advanced DNN-based systems to construct more accurate predictive models for breast cancer data. This method aids in the development of a precise and accurate clinical dataset classification tool for breast cancer. Furthermore, the model outperformed other well-known state-of-the-art algorithms and models in terms of architecture.

Andrea Apicella et al [5], have worked with a review on various modern trainable activation functions. The proposed functions are split into two main groups: fixed shape and trainable shape. For the second class of activation functions, two distinct groups of trainable activation functions have been found: parameterized standard functions and functions based on ensemble methods. The latter has been further enhanced by sorting out another family of activation functions, resulting in a linear combination of one-to-one functions.

Loris Nanni et al. [6], developed new models that may be utilized as standalone networks or as parts of an ensemble model. The base of this model approach to design is a way for matching particular components with the different functional units of the top-performing CNN models. The activation functions of each layer in the CNN model vary, and the total of the output tasks is evaluated. It emerged each activation layer in a CNN be replaced with an activation function chosen at random from a group of activation functions. A CNN with a unique set of activation function layers is a result of this.

Bawa V.S. et al. [7] proposed a variation of neural networks that can model non-linear dependencies in data. Even with a non-linear structure, the suggested activation function's behavior is still non-saturating. Various information range segments receive different activation behavior from the activation function. Two distinct iterations of the suggested function are shown in this study. The fixed structure activation function known as the "linear sigmoidal activation" function has function coefficients that are defined from the beginning of the model design process. On the other hand, the second function, known as the "adaptive linear sigmoidal activation" function, is trainable and can adjust to the complexity of the input data. The suggested models are evaluated using benchmark datasets (CIFAR-10, MNIST, SVHN, FER-2013) and state-of-the-art activation functions.

This article is organized as Section 3 includes Materials and Proposed Methodology, Section 4 presents results and discussion done thus far in the implementation on histology images of oral cancer, and Section 5 briefs the Conclusion and Future Work.

3. Materials and Proposed Methodology:

A. Dataset Description

Dataset collection for OSCC is the most important part of the research. The histopathologic images of OSCC are obtained from the Mendeley data [1], which has 1224 images with two different magnifications of 100x (10x objective lens \times 10x eyepiece) and 400x (40x objective lens \times 10x eyepiece) comprising 290 normal images and 934 cancerous images. A Leica ICC50 HD microscope was used to take pictures of H&E-stained tissue slides that were gathered, processed, and categorized by medical professionals from 230 patients.

B. Activation functions in the proposed CNN

The Convolutional Neural Network of our proposed model has an input layer, 3 hidden layers, and an output layer. Input is the folder of histopathological images. There are three convolutional layers with kernel sizes of (2x2) each followed by different activation functions (AF). These layers extract spatial features from the input image. Each convolutional block is followed by a max-pooling layer, reducing the spatial dimensions of the feature maps while retaining important features. After multiple convolution and pooling layers, the 3D feature maps are flattened into a 1D vector for the fully connected (dense) layers. The fully connected layers further process the flattened feature maps to make predictions. The network ends with a sigmoid activation function, commonly used for binary classification tasks.

The proposed CNN model classifies the dataset as benign or malignant cancerous images using disparate activation functions. To work with the proposed CNN model, the given dataset is insufficient. So various augmentation techniques including horizontal flipping, vertical flipping, rotation, shearing, hue, saturation, brightness, and contrast are used without affecting the original data. The activation function, which allows artificial neural networks to manage non-linear problems, is the most significant aspect of a network of neurons that decides whether a neuron will be stimulated and succeed to the next layer or not. Activation functions can be deployed to normalize the output so it falls between the interval 0 and 1 or -1 and 1. Various activation functions of both linear and nonlinear activation functions are used to measure the performance metrics of the given histopathological OSCC dataset image classification.

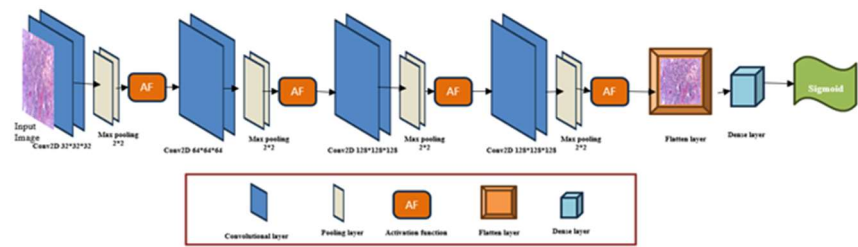
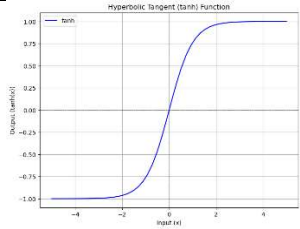
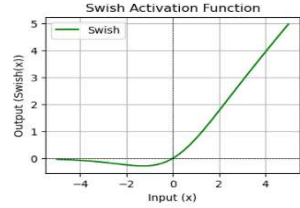
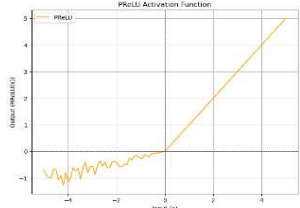
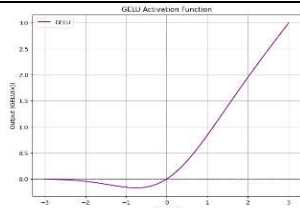
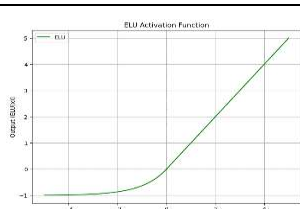


Fig 3: Proposed CNN architecture deploying disparate activation functions.

The use of different activation functions along with the graph representation, equation, derivation, and ranges are listed below in Table 1.

Table 1: Disparate activation functions with its graphical representation

Activation Function	Graph	Equation	Derivation	Range
ReLU		$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	$[0,\infty)$
Leaky ReLU		$f(x) = \begin{cases} 0.01 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0.01 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	$(-1,1)$

Tanh		$f(x) = \tanh(x)$ $= \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$f'(x) = 1 - f(x)^2$	$(-1,1)$
Swish		$f(x) = x \cdot \text{sigmoid}(x)$	$f'(x) = f(x) + \sigma(x)(1 - f(x))$	$(-\infty, \infty)$
PReLU		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$(-\infty, \infty)$
GeLU		$f(x) = xP(X \leq x)$ $= x\Phi(x)$	$f'(x) = \Phi(x) + x\Phi(x)$	$(-0.17..., \infty)$
ELU		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$	$f'(x) = \begin{cases} \alpha e^x & \text{if } x < 0 \\ 1 & \text{if } x > 0 \\ 1 & \text{if } x = 0 \text{ and } \alpha = 1 \end{cases}$	$[0, \infty)$

4. Results and Discussion

An automated histopathological OSCC classification is done using the proposed CNN model has been designed to avoid false diagnosis, time delay, and reduced human intervention. The proposed CNN in this study assessed how well different activation functions (CNNs) classified histopathological pictures. As these images are frequently distinguished by complex textures and variable staining patterns, the results show that the network's accuracy in classifying these images is highly dependent on the choice of activation function.

The loss and accuracy of the training OSCC dataset obtained are listed in Table 2. The loss and accuracy of the test OSCC dataset obtained are listed in Table 3. Of the given data sets, the Swish activation function proved to be more accurate, precise, and with minimal loss than the other investigated activation functions. Figures 4 and 5 show the analysis of the performance of the proposed CNN model for the training dataset and test dataset with various activation functions.

Fig. 6 and 7 show the confusion matrix of the ReLU activation function of the test and training dataset respectively. The confusion matrix for the Leaky ReLU activation function of the test and training datasets is displayed in Figures 8 and 9, respectively. The confusion matrix of both the test and training dataset. The Tanh activation function is highlighted in

Figures 10 and 11 respectively. Figures 12 and 13, respectively, demonstrate the confusion matrix of the Swish activation function across both the test and training datasets. In Figures 14 and 15, respectively, the PReLU activation function confusion matrix for the test and training datasets is highlighted. The GeLU activation function confusion matrix for test and training datasets is shown in Figures 16 and 17, respectively. The ELU Confusion Matrix for the training dataset and test dataset is presented in Figures 18 and 19, correspondingly.

Compared with the other activation functions, the training dataset, featuring 6901 cases of oral cancer and 6974 cases of noncancerous images, was precisely categorized by the proposed model using the swish activation function. The algorithm misclassified 99 cases of oral cancer images as non-cancerous and 26 instances of non-cancerous images as oral cancer images. The test dataset, which comprised 2987 instances of non-cancerous imagery and 2938 occurrences of oral cancer, was precisely identified by the proposed model using the swish activation function. The model incorrectly categorized 62 cases of oral cancer visuals as noncancerous and 13 cases of noncancerous images as oral cancer images. The proposed model has a 99% training accuracy and a 98% test accuracy when using the Swish activation function. The proposed model's training and test losses with the swish activation function are 0.025 and 0.040, respectively. Overall, this work satisfies the goal of enhancing automated diagnostic tools in medical imaging by offering insightful information about CNN architecture optimization for histopathological image processing.

Table 2: Accuracy and loss of **training** OSCC dataset using disparate activation functions

Activation Function	Training Accuracy	Training Loss
ReLU	50 %	0.69
Leaky ReLU	91 %	0.22
Tanh	50 %	0.70
Swish	99 %	0.025
PReLU	98 %	0.63
GeLU	98 %	0.05
ELU	69%	0.5

Table 3: Accuracy and loss of **test** OSCC dataset using disparate activation functions

Activation Function	Test Accuracy	Test Loss
ReLU	50 %	0.69
Leaky ReLU	89 %	0.27
Tanh	50 %	0.70
Swish	98 %	0.040
PReLU	97 %	0.09
GeLU	97 %	0.05
ELU	50%	0.69

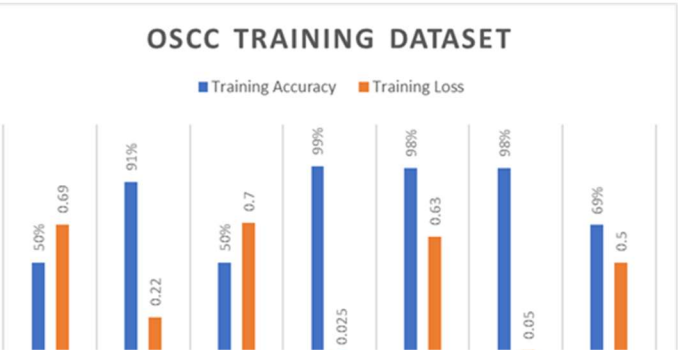


Fig. 4: Proposed CNN model performance analysis with training dataset using disparate activation functions

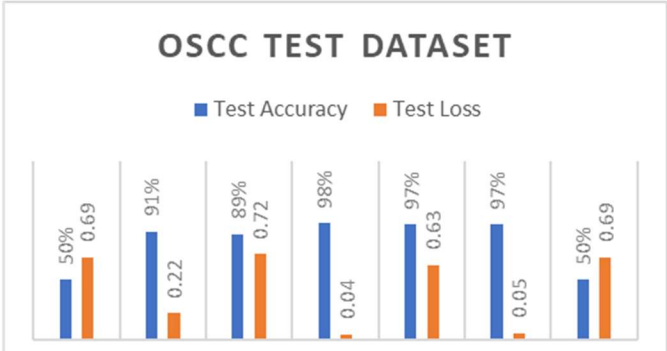


Fig. 5: Proposed CNN model performance analysis with test dataset using disparate activation functions

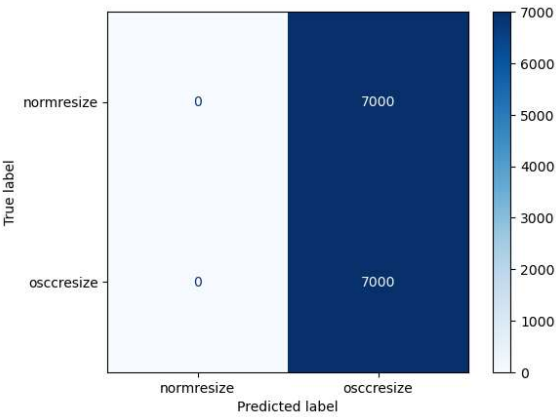


Fig.6: ReLU Confusion matrix for training dataset

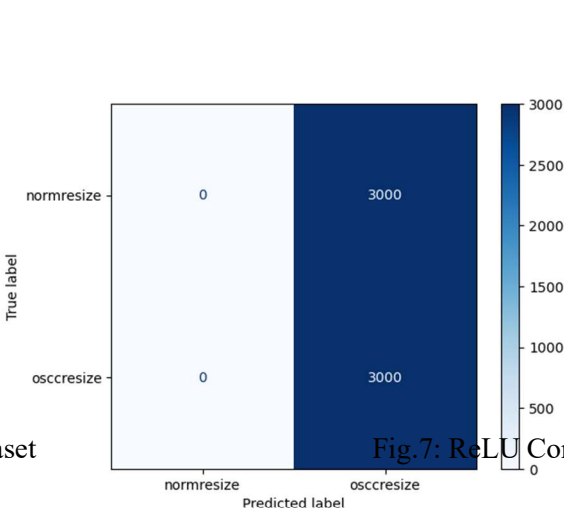


Fig.7: ReLU Confusion matrix for test dataset

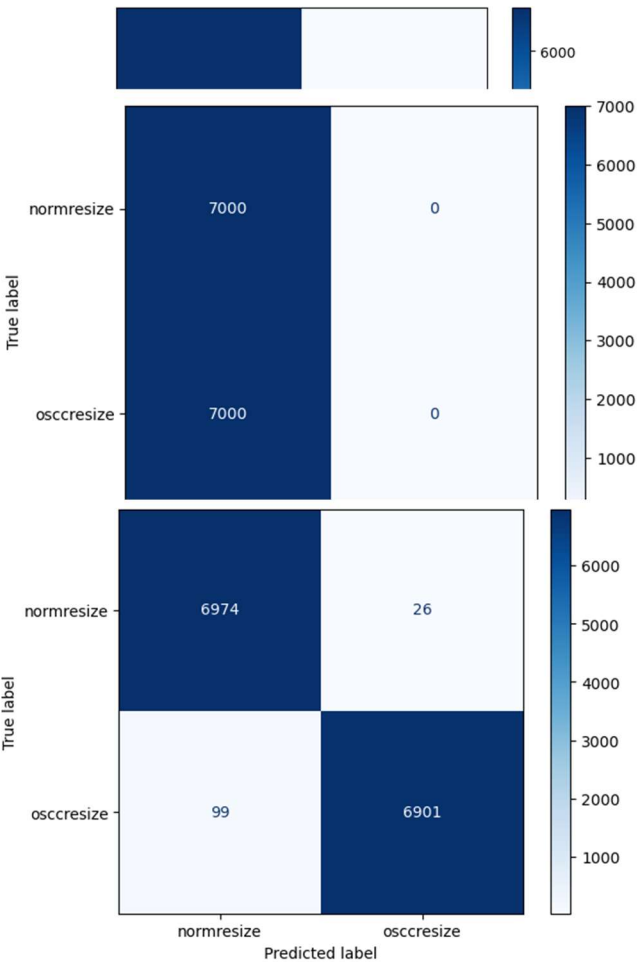


Fig.8: Leaky ReLU Confusion matrix for training dataset

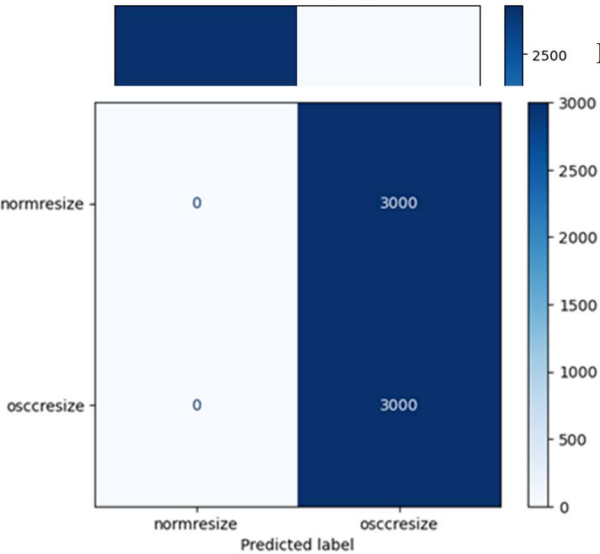


Fig.9: Leaky ReLU Confusion matrix for test dataset

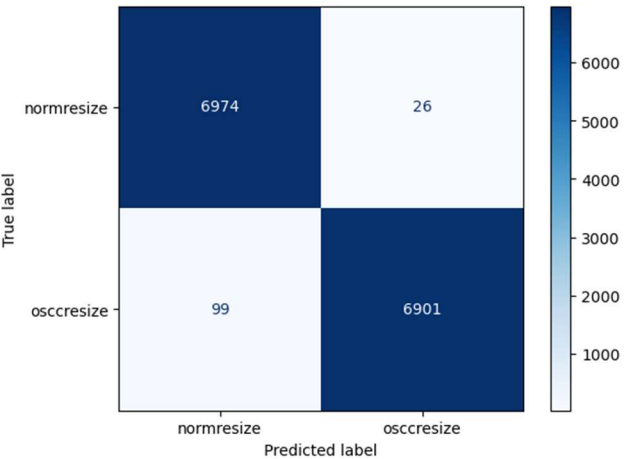


Fig.10: Tanh Confusion matrix for training dataset

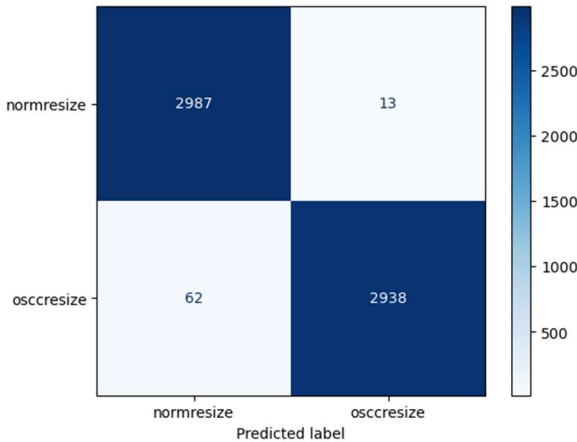


Fig.11: Tanh Confusion matrix for test dataset

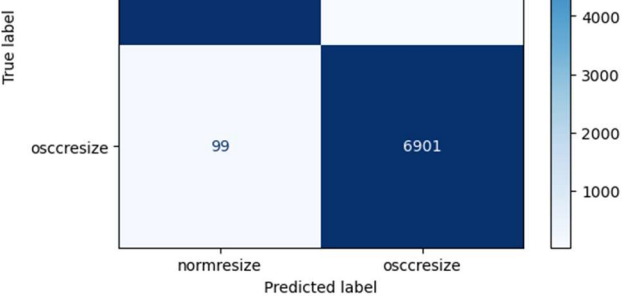


Fig.12: Swish Confusion matrix for training dataset

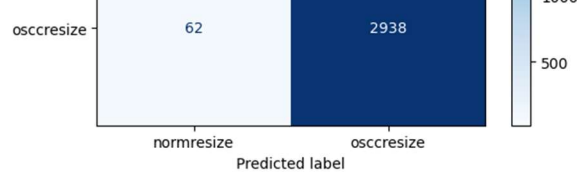
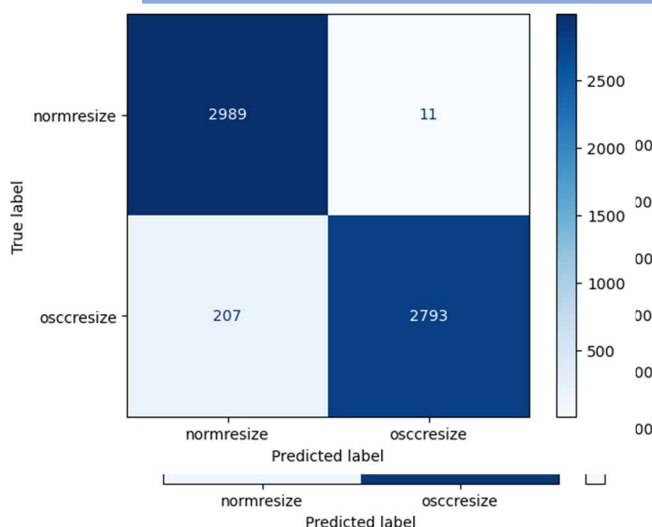


Fig.13: Swish Confusion matrix for test dataset

matrix for test dataset



Confusion matrix for training dataset

Fig.16: GeLU Confusion matrix for training dataset
for test dataset

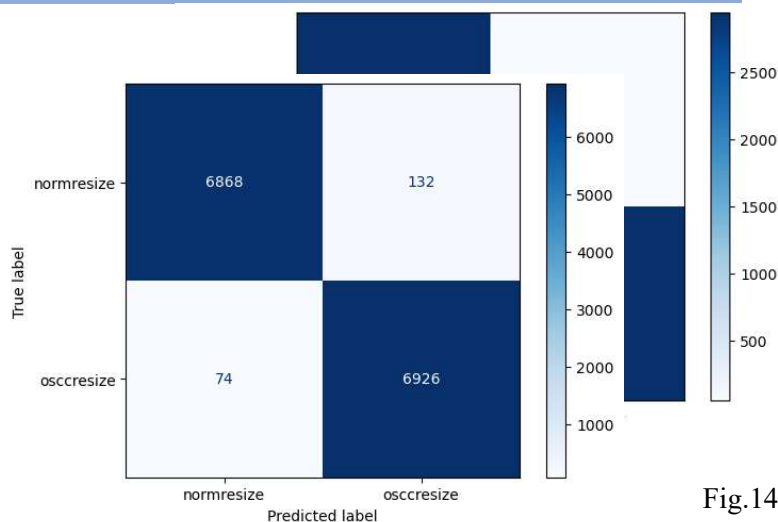


Fig.14:
PRelu

Fig.15: PReLU Confusion matrix for test dataset

Fig.17: GeLU Confusion matrix

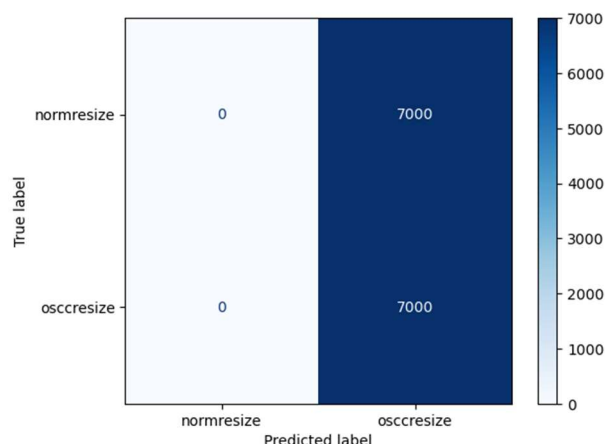


Fig.18: ELU Confusion matrix for training dataset
matrix for test dataset

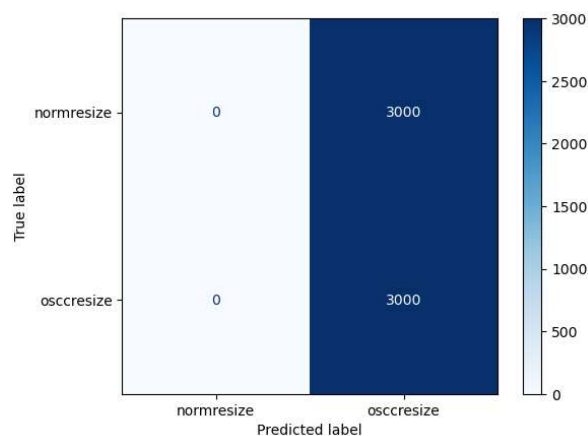


Fig.19: ELU Confusion

5. Conclusion and Future Work

The proposed CNN model imparts the importance of the activation functions that play a major role in the hidden layer in classifying the histopathological image dataset. The Swish function performed well with minimal loss function, and maximum accuracy was obtained from these activation functions. Images with different magnifications of 100x of low power magnification and 400x of high-power magnification are used to perform image classification accurately. Morphological extraction is to be done to find out the exact position of the tumor cells. In future work, these activation functions to be applied to various machine learning and deep learning predefined models with the annotation

of nuclei in OSCC histology images. The ensemble model will be developed using traditional models to perform morphological operations and texture analysis.

Acknowledgment

This work was Funded by MHRD RUSA 2.0 “Biomaterials” Functional Synthetic Materials for Biomedical Applications (Theme-2) - Nanoinformatics on the role of Nano-HAp in cancer imaging, orthodontics and dental: Schema and System Design. Our sincere gratitude to Dr. Prabhakaran Mathialagan, Jain (Deemed-To- Be) University, Bangalore for his support and guidance.

References

1. Rahman, Tabassum Yesmin (2019), “A histopathological image repository of normal epithelium of Oral Cavity and Oral Squamous Cell Carcinoma”, Mendeley Data, V1, doi: 10.17632/ftmp4cvtmb.1
2. Nanni L, Brahnam S, Paci M, Ghidoni S. Comparison of Different Convolutional Neural Network Activation Functions and Methods for Building Ensembles for Small to Midsize Medical Data Sets. *Sensors*. 2022; 22(16):6129. <https://doi.org/10.3390/s22166129>
3. Wong, K., Dornberger, R. & Hanne, T. An analysis of weight initialization methods in connection with different activation functions for feedforward neural networks. *Evol. Intel.* (2022). <https://doi.org/10.1007/s12065-022-00795-y>
4. Vijayakumar K, Kadam VJ, Sharma SK. Breast cancer diagnosis using multiple activation deep neural network. *Concurrent Engineering*. 2021;29(3):275-284. doi:10.1177/1063293X211025105
5. Apicella A, Donnarumma F, Isgrò F, Prevete R. A survey on modern trainable activation functions. *Neural Netw.* 2021 Jun;138:14-32. doi: 10.1016/j.neunet.2021.01.026. Epub 2021 Feb 9. PMID: 33611065.
6. Nanni L, Lumini A, Ghidoni S, Maguolo G. Stochastic Selection of Activation Layers for Convolutional Neural Networks. *Sensors*. 2020; 20(6):1626. <https://doi.org/10.3390/s20061626>
7. Vivek Singh Bawa, Vinay Kumar, Linearized sigmoidal activation: A novel activation function with tractable non-linear characteristics to boost representation capability, *Expert Systems with Applications*, Volume 120, 2019, Pages 346-356, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2018.11.042>.
8. M. Du, R. Nair, L. Jamieson, Z. Liu, P. Bi, Incidence trends of lip, oral cavity, and pharyngeal cancers: global burden of disease 1990–2017, *Journal of dental research* (2) (2020) 143–151.
9. S. Perdomo, G. M. Roa, P. Brennan, D. Forman, M. S. Sierra, Head and neck cancer burden and preventive measures in central and South America, *Cancer epidemiology* (2016) S43–S52.
10. K. Joarder and S.M.Aziz, “A note on activation function in multilayer feedforward learning”, *Proceedings of International Joint Conference on Neural Networks: IJCNN ‘02*. Vol. 1, pp 519-523, 2002.
11. Rawat W, Wang Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computing*. 2017 Sep;29(9):2352-2449. doi: 10.1162/NECO_a_00990. Epub 2017 Jun 9. PMID: 28599112.
12. Almeida MAM, Santos IAX. Classification Models for Skin Tumor Detection Using Texture Analysis in Medical Images. *Journal of Imaging*. 2020; 6(6):51. <https://doi.org/10.3390/jimaging6060051>.
13. Shiv Ram Dubey, Satish Kumar Singh, Bidyut Baran Chaudhuri, Activation functions in deep learning: A comprehensive survey and benchmark, *Neurocomputing*, Volume 503, 2022, Pages 92-108, ISSN 0925-2312, <https://doi.org/10.1016/j.neucom.2022.06.111>.
14. Chu CS, Lee NP, Adeoye J, Thomson P, Choi SW. Machine learning and treatment outcome prediction for oral cancer. *Journal of Oral Pathology & Medicine*. 2020 Nov; 49(10):977-85.
15. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021;8(1):53. doi: 10.1186/s40537-021-00444-8. Epub 2021 Mar 31. PMID: 33816053; PMCID: PMC8010506.
16. P. Mathialagan and M. Chidambaranathan, "Morphological and Characteristic Analysis of Upper Aero-Digestive Tract Tumour: Revealing Uncovered Facts in Digital Pathology," 2021 International Conference on Computing,

Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2021, pp. 550-556, doi: 10.1109/ICCCIS51004.2021.9397133.