IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 1, March 2024, pp. 487~499
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i1.pp487-499

An efficient convolutional neural network-based classifier for an imbalanced oral squamous carcinoma cell dataset

Usha Manasi Mohapatra[1], Sushreeta Tripathy[2]
[1]Department of Computer Science, Gangadhar Meher University, Sambalpur, Odisha, India
[2]Department of Computer Application, S'O'A Deemed to be University, Bhubaneswar, Odisha, India

Article Info

Article history:

Received Jan 25, 2023
Revised May 4, 2023
Accepted May 7, 2023

Keywords:

Class imbalance
Convolutional neural network
Medical images
Oral squamous cell carcinoma
Oversampling

ABSTRACT

Imbalanced datasets pose a major challenge for the researchers while addressing machine learning tasks. In these types of datasets, samples of different classes are not in equal proportion rather the gap between the numbers of individual class samples is significantly large. Classification models perform better for datasets having equal proportion of data tuples in both the classes. But, in reality, the medical image datasets are skewed and hence are not always suitable for a model to achieve improved classification performance. Therefore, various techniques have been suggested in the literature to overcome this challenge. This paper applies oversampling technique on an imbalanced dataset and focuses on a customized convolutional neural network model that classifies the images into two categories: diseased and non-diseased. Outcome of the proposed model can assist the health experts in the detection of oral cancer. The proposed model exhibits 99% accuracy after data augmentation. Performance metrics such as precision, recall and F1-score values are very close to 1. In addition, statistical test is performed to validate the statistical significance of the model. It has been found that the proposed model is an optimised classifier in terms of number of network layers and number of neurons.

Corresponding Author:

Usha Manasi Mohapatra
Department of Computer Science, Gangadhar Meher University
Sambalpur, Odisha, India
Email: ushamanasimohapatra@gmail.com

1. INTRODUCTION

With growing availability of large scale of unstructured and complex data required for prediction and classification functions, it has been a critical task to extract summarised information to support decision making. Data analysing tools and knowledge discovery techniques have exhibited tremendous success in several real world applications such as recommendation systems, financial market analysis, customer review analysis and many more. Despite the success history, some data groups fail to address the predictive analytical problems.

One of the reasons behind such failures for decision making is the class imbalance dataset. The model which is trained for such data is tuned more towards the majority samples. Hence, processing such skewed data often produces biased results. It has been reported in the literature [1], [2] as a crucial factor in training the imbalanced data. Most classifiers assume equal distribution of individual class instances. Hence, when these algorithms are presented with imbalanced datasets, they lack generalization and exhibit poor performance metrics. Past studies highlight the implications of binary imbalanced datasets in biomedical applications [3]. Most often, real time data collected in the health sector suffer from such a problem. Due to the significant difference in number of instances of individual classes, machine learning (ML) algorithms tend to exhibit inappropriate results [4]. Sometimes, the performance measures of the classifiers guide towards misleading conclusions out of the model behaviour. For example, consider a dataset with class distribution as 20%:80%. It means that for one class (positive class) the number of sample instances is 80 and that is 20 for the other class (negative class). Even if the classification model results into 90% accuracy, the model won't be considered good because the negative class instances are projected as positive that enhances the false positive metric of the model. Though logical, it is an undesired consequence [5], [6].

Skewness in class samples is also very pervasive in many data mining applications namely text classification [7], risk management, detection of oil spills in satellite radar images of ocean surfaces, medical diagnosis, the detection of fraudulent calls, and spam mail recognition. Class imbalance problems are addressed by many techniques out of which two ways are mostly reported in literature [8]. One is to undersample the majority class instances [9], [10] and the other one is to generate synthetic data from minority class tuples. In [9], a technique synthetic minority oversampling technique (SMOTE) is proposed that generates new samples from existing samples of minority class; i) The major contributions of the research article are as follows; ii) Employ oversampling to reduce the difference in class frequencies of data samples; iii) Set up a model by properly setting the hyperparameters for effective binary image classification; iv) Evaluate the model using performance measures like precision, recall, and area under curve; v) Apply the model for two different imbalanced medical image datasets; vi) To confirm the statistical significance of the classification model using McNemar test.

Remaining part of the paper is comprised of six more sections. Section 2 describes related work collected from existing literature. Objectives of the work are stated in section 3. Section 4 deals with basics of convolutional neural network (CNN) and proposed methodology. Data collection and processing are presented in section 5. Results and discussions are elaborated in section 6. At last, section 7 concludes the study with possible future scope.

2. RELATED WORK

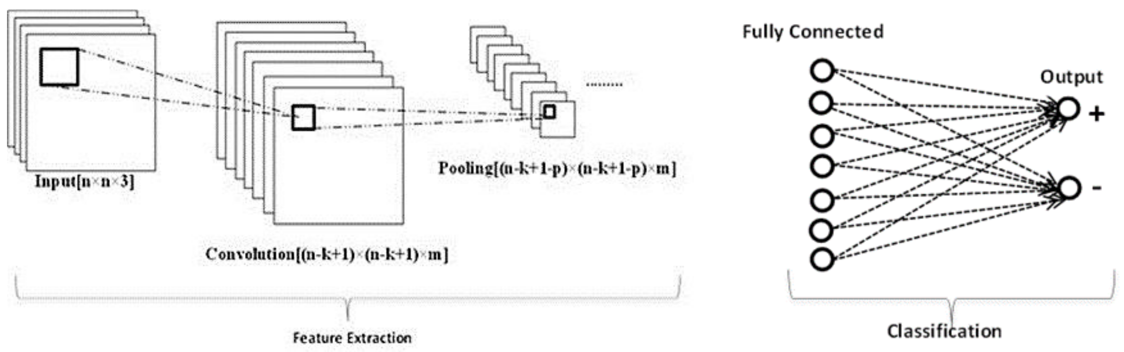
For this study, different research article databases namely Science Direct, IEEE Xplore, Springer and Web of Science have been searched. Specifically, browsing is based on keywords like 'Classification for oral squamous cell carcinoma (OSCC) dataset', 'Data augmentation for image', and 'machine learning for imbalanded image dataset'. Current study focuses on recently published research articles based upon machine learning algorithms for imbalanced medical image datasets. Other cited documents has been referred to discuss the efficiency of machine learning tools in various domains, performance measures of the classifiers, and data sampling applications for imbalanced datasets. Summary of all the referred papers that employ some form of deep learning methods for imbalanced datasets is framed in Table 1 [1]-[35] (see in Appendix). It provides the literature summary table that includes the synopsis of all the related works considered in this study.

3. OBJECTIVES

The summary table of related works point out the application of several deep learning and data augmentation techniques adopted for imbalanced medical image datasets. However, the efficiency of those models is bounded upto 92% in terms of F1 score and 95% in terms of area under curve (A_UC) respectively. The main objective of our study is to minimize the failure rate in classification for class imbalance dataset. By inspctful hyperparameter tuning, the proposed binary classifier reduces both false positive and false negative rate to nearly 0. In this work, a customized convolutional neural network is presented to classify OSCC images with 99% accuracy. The performance of the model is confirmed against a statistical McNemar's Test. Data collected for the study suffers from the disproportionate class sample distribution problem which has been overcome by data augmentation techniques, the outcome of the proposed model may assist the health experts in the detection of oral squamous cell carcinoma. The proposed model exhibits promising classification results compared to the existing state of the art models.

4. METHODOLOGY

Advancements in computing power and algorithm efforts have led to the tremendous ability of deep learning techniques in analysing medical images [25]–[28]. These computer assisted findings can be used as an alternative cross verification tool for pathology tests by healthcare professionals. Deep learning [30], [31] methods have been adopted in different domains for the task of object detection, image segmentation, image classification and so on. In contrast to traditional machine learning algorithms in which features are extracted computationally, CNN helps the data analyst by automatically drawing out those. Nevertheless, feature map is also reduced significantly. The standard process of CNN is depicted in Figure 1(a) that shows only feature extraction part. Figure 1(b) displays the classification segment. Any CNN model takes input as the matrix of numbers. This paper employs images as the input to proposed CNN architecture. Each image is represented as a matrix of numbers. So, if one image is given as input to the network, then the input matrix consists of n×n×3 numbers where 3 represents number of channels in the image. The convolution operation is applied to the image along with kernel matrices to extract features. As mentioned in Figure 1(a), the number of pixels after convolution is reduced from n×n×3 to (n-k+1) × (n-k+1) ×3, where k×k is the size of the kernel. The convoluted image is passed through some sort of pooling to reduce the dimension of the feature space. The pixel size in the image after pooling is further reduced to (n-k+1-p) ×(n-k+1-p)×3, where p×p is the pooling dimension. An image may undergo a sequence of convolution and pooling repeatedly depending on the problem and data. CNN differs from artifcial neural networks (ANN) upto this phase. Once desired number of application of convolution and pooling completes], the pixels are flattened into one dimensional array. This one dimensional array is fed to the ANN for classification as demonstrated in Figure 1(b) or any other data mining task. The workflow of the proposed model is presented in Figure 2.



(a) (b)

Figure 1. CNN architecture; (a) feature extraxtion in CNN and (b) classisication using ANN

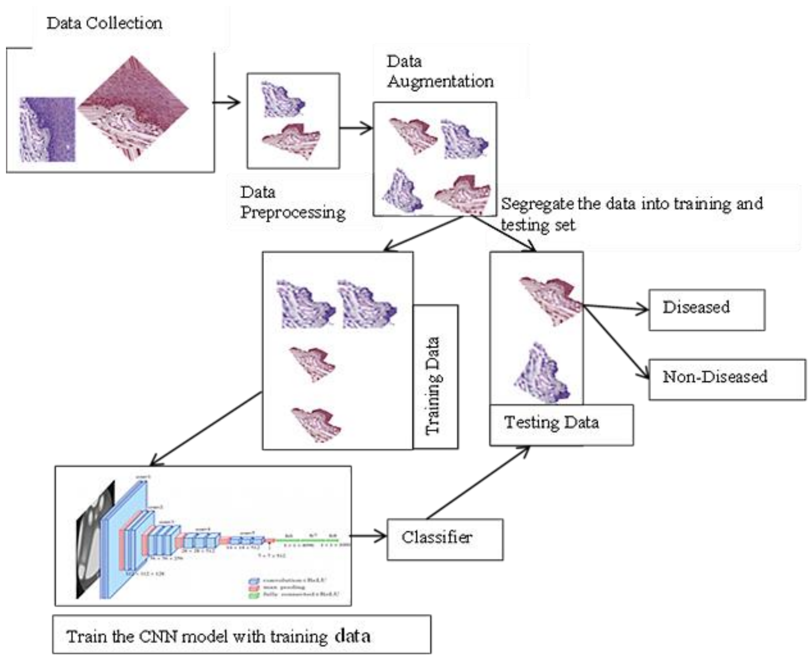


Figure 2. Work flow of the proposed model

4.1. Work flow of the model

In this research paper, the authors have experimented on a simple architecture of CNN which is a sequential model. It consists of two layers: each layer includes convolution and pooling. First layer convolution takes 32 filters having kernel size 3×3. Max pooling has been chosen in pooling step to select the most contributing features of the image which are the pixels with high intensities.

The second layer comprises of convolution with 64 filters. In both of these layers pool size has been taken as 2×2 and activation function is rectified linear unit (ReLU). In Algorithm 1, the procedure for predicting the disease of oral cancer is presented.

Algorithm 1: Oral cancer prediction technique using CNN

```
Input: Labelled Image Dataset
Output: Diseased / Non-diseased
Begin
    1.      Augment the images
           /*so that the gap in the number of positive and negative image samples
        becomes
            negligible*/
    2.      Resize the images of size (x*y*z) to (x'*y'*z) where x' < x
    3.      Split the dataset into two parts: Training (80%) and Testing (20%)
    4.      Normalize both the set of images
    5.      Initialize the parameters of the proposed model
    6.      Train the model with training data
    7.      If (model performance is acceptable) then
            Test the model with Testing data
    8.      else
            reconfigure the model, goto step 6
    9.      Then introduce unseen image data to the model for classification of the
        image either
            into Diseased class or non-diseased class
```

5. DATA COLLECTION AND PRE-PROCESSING

The dataset used for implementation of the proposed model is referred from [32]. The online repository consists of two datasets of images in .jpg format. Images of two different resolutions are organised into two different directories. One directory consists of histopathological images with the normal epithelium of the oral cavity in 89 numbers and 439 images of oral squamous cell carcinoma (OSCC) in 100x magnification.

Another directory consists of 201 images with the normal epithelium of the oral cavity and 495 histopathological images of OSCC in 400x magnification. Some samples of both normal and carcinoma affected images are shown in Figure 3 and Figure 4 respectively. The ratio of the normal image to the OSCC image in first directory is 1:20 which means the data is skewed. Hence, oversampling of minor class data sample has been adopted to get the dataset appropriate for the model. The image augmentation technique is followed for the implementation of image oversampling.

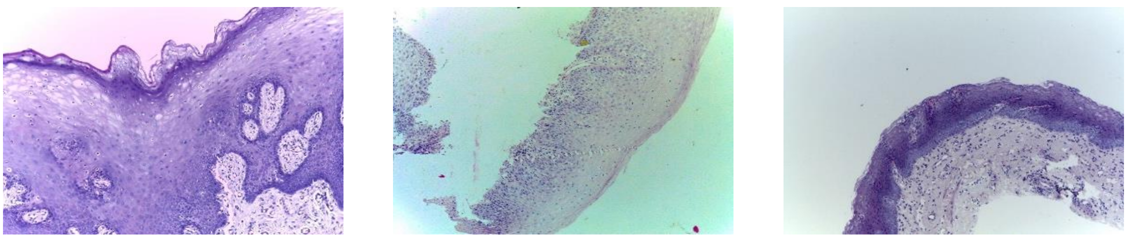


Figure 3. Samples of normal images

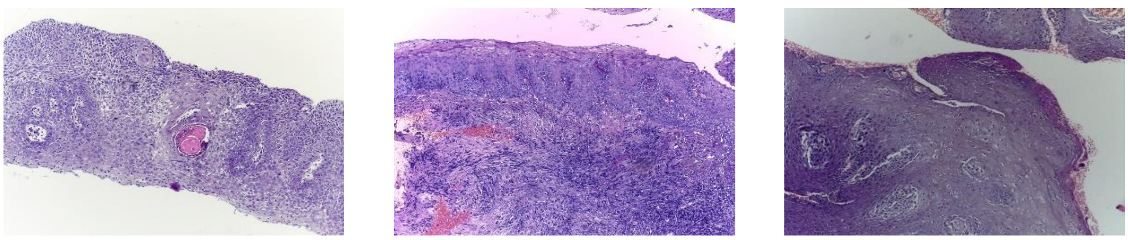


Figure 4. Samples of OSCC images

In image augmentation methodology, each image undergoes different transformations like rotation, shifting, shearing, flipping, zooming and modifying other properties like brightness. For the training set, authors have synthesized many transformed images from each original image by using ImageDataGenerator class available in the python library. A sample snapshot of derived images is shown in Figure 5. Once the numbers of image samples in normal and carcinoma categories are closely equal, further pre-processing is applied to the data.

Pre-processing is required to remove noise and outliers, handle missing values, bring the data into same scale, and normalization. The current study has only implemented max normalization to transform the data into a range (0, 1). Then it is split into training and testing set in the ratio 80:20. Number of images in training set is 700 whereas that in testing is 176. The training data is convolved with 32 fiters of dimension 3x3. The experimental setup for CNN has considered ReLU as the activation function.

Usually, amongst many activation functions, researchers prefer ReLU because it does not perform expensive computations and in practice, shows better convergence performance. After the CNN operations, the processed image pixels are flattened and fed as the input to the ANN layer. Finally in the output layer, the sigmoid activation function is used to classify the image into either normal class or carcinoma class.

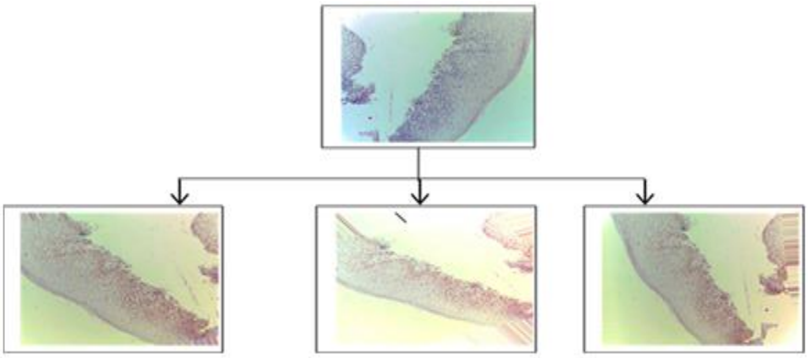


Figure 5. Samples of augmented images

6. RESULTS AND DISCUSSIONS

This section presents the results obtained from simulation and their interpretations. Literature on data mining mention different performance metrics of classifiers like accuracy, sensitivity(recall), specificity, precision, F1-score, confusion matrix, receiver operating characteristic (R_OC)-A_UC, Log-loss and so on. Mathematical formula for some of these performance metrics are provided in (1) to (7). Some common terminologies used in classification are listed in Table 2. Depending on the problem statement, the meaning of positive is decided. For example, in the given problem, detection of carcinoma cells is considered as positive.

T_P means an image is originally has carcinoma and is predicted as also carcinoma. Similarly, if an image does not have carcinoma and is predicted as non-carcinoma, then it is treated as T_N. On the contrary, if an image has carcinoma but is not predicted as carcinomous, then it is considered as F_{N0}. F_P means the actual image does not have carcinoma but predicted as carcinomous. Hence, accuracy is computed as the ratio of total number of correct predictions and total number of predictions.

$$Accuracy = \frac{T_P+T_N}{T_P+T_N+F_P+F_N} \tag{1}$$

Table 2. List of acronyms used for classification performance metrics

Acronym	Full Form
T_P	True Positive
F_P	False Positive (Type 1 Error)
T_N	True Negative
F_N	False Negative (Type 2 Error)
T_PR	True Positive Rate
F_PR	False Positive Rate
T_NR	True Negative Rate
F_NR	False Negative Rate
R_OC	Receiver Operating Characteristic
A_UC	Area Under Curve

But, accuracy is not an appropriate performance indicator for a model trained with imbalanced data. Precision is another metric which determines out of all predicted positive cases how many are actually positive. It is useful in problems where F_P cases are to be reduced.

$$Precision = \frac{T_P}{T_P+F_P} \quad (2)$$

The recall does not include information about the F_P cases. It only finds the ratio of T_P and total number of actual positives. It indicates how good the model is in detecting all the T_P cases. It is also referred as sensitivity (same as T_PR). Specificity (same as F_PR) is defined as the proportion of actual negatives, which got predicted as the negative.

$$Recall = \frac{T_P}{T_P+F_N} \quad (3)$$

$$Specificity = \frac{T_N}{T_N+F_P} \quad (4)$$

$$F1score = \frac{2*Precision*Recall}{Precision+Recall} \quad (5)$$

$$F_PR = \frac{F_P}{T_N+F_P} \quad (6)$$

$$F_NR = \frac{F_N}{T_P+F_N} \quad (7)$$

F_1 score is the harmonic mean of precision and recall. The confusion matrix is a two-dimensional array in which the cells indicate T_P, T_N, F_P, and F_N cases. It helps in estimating what way the model is correct or wrong. The confusion matrix for binary class problem is depicted in Table 3.

Table 3. Confusion matrix for binary class problem

	Predicted Negative	Predicted Positive
Actual Negative	True Negative (T_N)	False Positive (F_P)
Actual Positive	False Negative (F_N)	True Positive (T_P)

Another important metric is the R_OC curve. Basically, it is used for inspecting the output quality of a binary classifier at different threshold settings. This curve is plotted against two parameters: T_PR (shown in Y-axis) and F_PR (shown in X-axis). In some literature, it is also suggested to take other parameters along X-axis. An example of R_OC curves is shown in Figure 6.

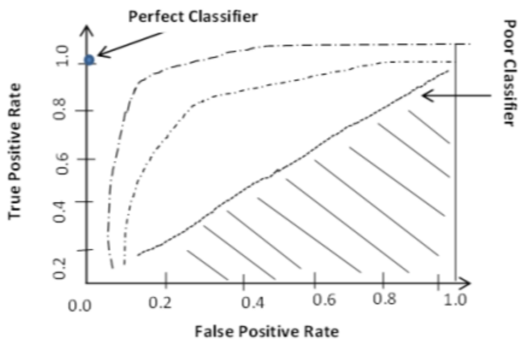
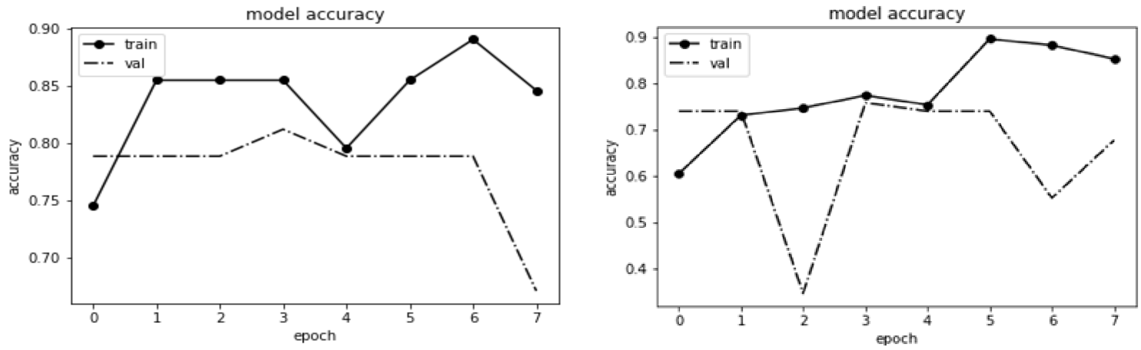


Figure 6. Demonstration of R_OC curves

It is a curve that is plotted taking different pairs of (T_PR, F_PR) values for different classification thresholds. In the Figure 3 different ROC curves are plotted. Each curve corresponds to one classifier. The area under aparticular R_OC curve is termed as A_UC. It is marked with slanted lines in the Figure 6. The classifier whose R_OC curve covers a large area is considered as a better classifier. As per this convention, the classifier corresponding to top R_OC curve can be acknowledged as best among three classifiers shown in figure. The current study has undertaken two datasets each consisting of unequal proportion of non-diseased (normal) and diseased (carcinoma) images. Description of the dataset is provided in Table 4. Due to the skewness present in the data, the proposed model is an overfitted model which can be affirmed from the plot of validation data accuracy and loss during the training phase of the model. The plots for two different unaugmented datasets are provided in Figures 7(a)-7(b) and Figures 8(a)-8(b).

Table 4. Description of dataset

	No.of Normal Images	No.of Carcinoma Images
Dataset 1	89	439
Dataset 2	201	495

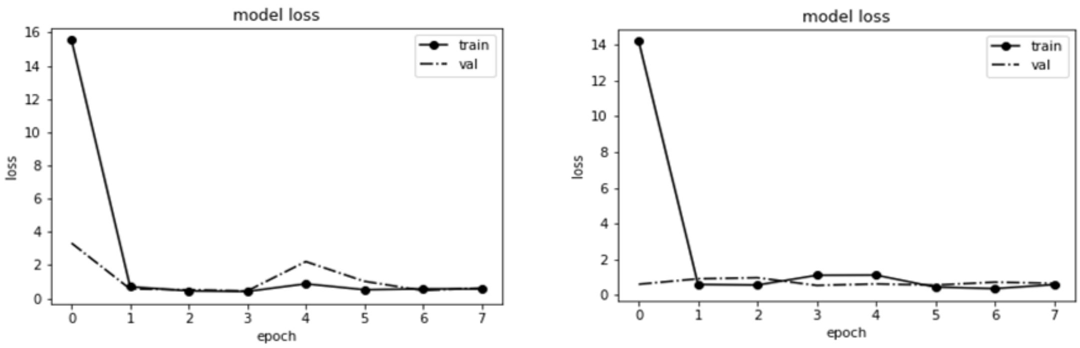


(a) (b)

Figure 7. Accuracy plot during training for un-augmented data (a) dataset 1 (b) dataset 2

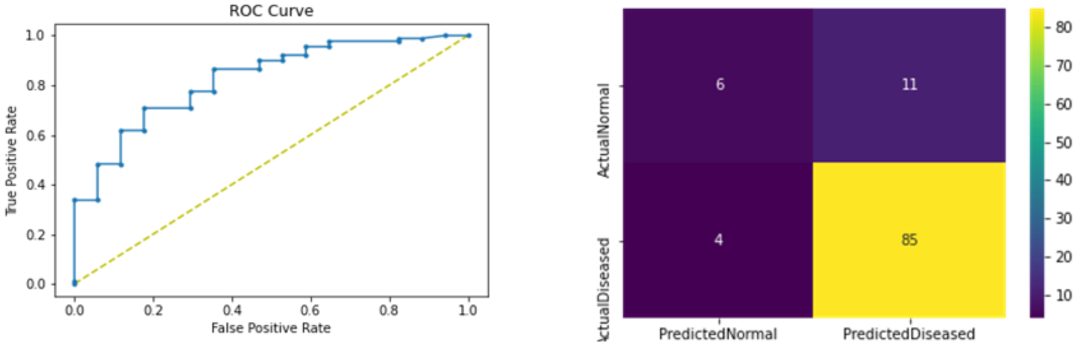
As observed from the figure, there is a huge gap between training accuracy compared to validation accuracy though the rate of loss in training data is satisfactory and accuracy is nearly 0.9. Due to the presence of data imabalance, the classification performance is poor in terms of F_PR. The poor classification performance can be observed from the R_OC plot which is depicted in Figure 9(a). The confusion matrix representing the F_P, T_P, F_N, and T_N is presented in the form of a heatmap in Figure 9(b).

To overcome the model overfitting, an augmentation technique has been applied using ImageDataGenerator class available in Python library. The class expands the datasets by transforming images through various transformation techniques. After the image augmentation, the model could classify the subjects with full accuracy that can be validated from Figure 10. Due to data augmentation technique, the proposed model is trained with a balanced dataset which removed the biased outcome of the model. Figure 10(a) demonstrates the R_OC plot drawn for the augmented dataset 1 using F_PR and T_PR of the proposed classifier. The confusion matrix generated for the classifier is depicted in Figure 10(b).



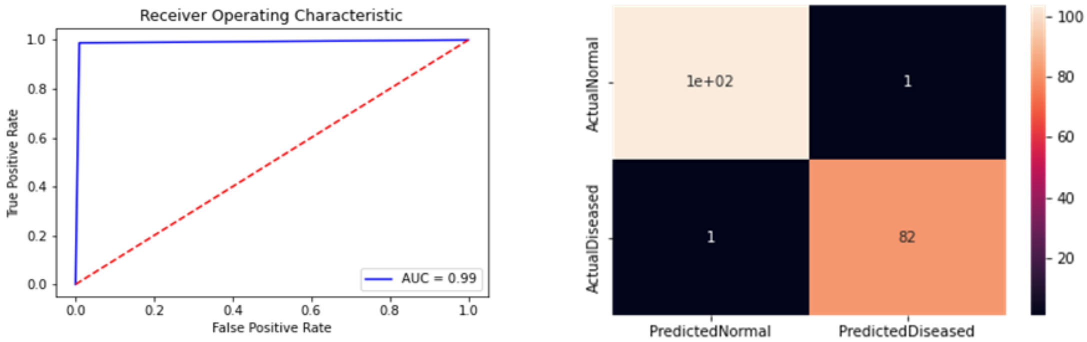
(a) (b)

Figure 8. Loss Plot during training for Un-augmented data; (a) dataset 1 and (b) dataset 2



(a) (b)

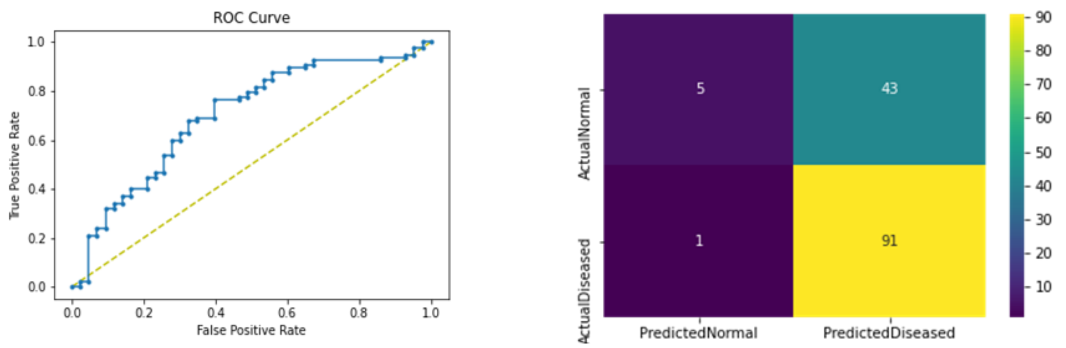
Figure 9. Performance measures for un-augmented dataset 1; (a) R_OC and (b) confusion matrix



(a) (b)

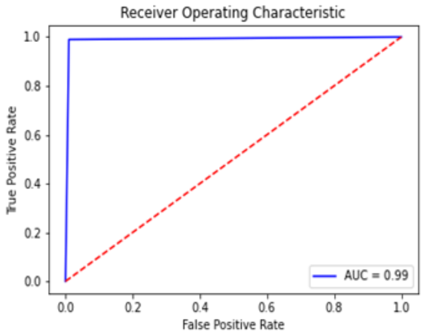
Figure 10. Performance measures for augmented dataset 1; (a) R_OC and (b) confusion matrix

It can be seen that the A_UC for the R_OC is ideal compared to that of the un-augmented data. The confusion matrix for augmented dataset 1 is outlined in Figure 10(b). Similarly, for un-augmented dataset 2, the R_OC and confusion matrix are portrayed in Figures 11(a) and 11(b) respectively. R_OC and confusion matrix for the augmented dataset 2 has been delineated in Figures 12(a) and 12(b) respectively. Furthermore, the authors depict the results through bar plots for clarity in visualization. Figure 13(a) showcases Accuracy, F_PR, and F_NR comparison for both unaugmented and augmented data considered in dataset 1. Similarly, Figure 13(b) illustrates Accuracy, F_PR, and F_NR comparison for both unaugmented and augmented data considered in dataset 2. It can be observed from both plots that even if there is a thin difference between corresponding accuracies, a remarkable gap is present in F_PR and F_NR. In the healthcare decision-making process, false predictions are very hazardous.

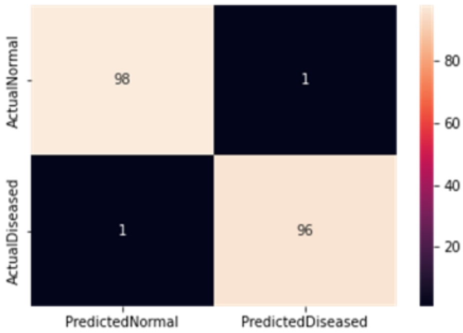


(a) (b)

Figure 11. Performance measures for un-augmented dataset 2; (a) R_OC and (b) confusion matrix

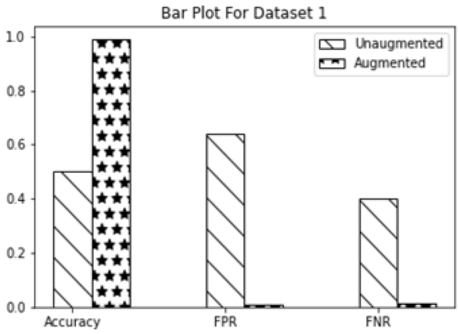


(a)

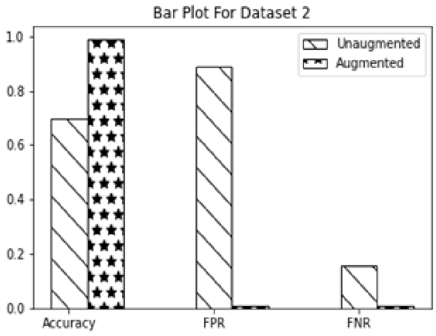


(b)

Figure 12. Performance measures for augmented dataset 2; (a) R_OC and (b) confusion matrix



(a)



(b)

Figure 13. Bar plot demonstrating accuracy, F_PR and F_NR; (a) for dataset 1 and (b) for dataset 2

The statistical significance of the proposed model is confirmed by McNemar's test. This test is applied on 2×2 contigency table. In this article the contigency table is the confusion matrix which stores the discordant pairs. In health domain, the rate of false predictions is as decisive as of true predictions. A model may not be considered as a worth one even if it gives 90% correct predictions as because, a significant number of false predictions could be fatal. McNemar's test is applied to determine the probability of difference between false positive and false negative predictions. The chi-square distribution is used to compute whether the row and column marginal frequencies are equal for paired samples. For the test, null (H0) and alternate (H1) hypotheses are defibed as follows:

H0: There is no significant difference between the marginal proportions of discordant pairs.

H1: There is a significant difference between the marginal proportions of discordant pairs.

McNemar's test statistic is computed as follows:

$\chi^2 = \frac{(b-c)^2}{b+c}$, where b and c are the discordant pairs from the confusion matrix. The degree of freedom is computed as (2-1)*(2-1)=1. The study considers 5% significance level for the test after reference to the literature. P-value for the test is obtained as 0.012 which is less than 0.05. Hence, the null hypotheses is rejected. The inference from the statistical test is there is marginal differences between the discordant pairs.

7. CONCLUSION

The proposed classifier implements a convolutional neural network to categorise medical images into either of the two classes: diseased and normal. The output of the classifier will extend an additional affirmation regarding the presence of carcinoma cells in the oral cavity. Datasets containing the images are skewed because the number of samples of diseased subjects is in multiples of that of normal subjects. Hence, the performance of the classifier was impoverished which is demonstrated through ROC and confusion matrix plots. To bring diversity and quality into the data different transformations such as rotation, scaling, shifting and flipping are applied to oversample the minority class instances. After application of the transformations, modified datasets are again applied to the model for training which enhances the classification performance to the superiority. Nevertheless, the suggested CNN model involves less complexity and time efficient as only two layers of convolution and pooling have been employed before flattening the input. This model also exterminates the use of cloud services as software as a service. The study has not performed any experiment on the augmentation technique for multiclass/non-binary class problems. This objective may be explored in the future extension of the current research work.

APPENDIX

Table 1. Summary table of related work

Sr#	Cite	Methodology	Dataset	Findings	Limitations
1	[1]–[3]	Review on challenges in data mining	Multi domain data	Problems in data mining task	Limited data source
2	[4]	Neural network with back propagation, Particle swarm optimization, Hybrid sampling	Breast cancer, Simulated data	Class imbalanced datasets have detrimental effect on classifier performance irrespective of the correlation factor and number of features	Study has not caliberated the classification performance measures like precision and recall
3	[5], [6]	Deep network	Magnetic resonance imaging (MRI), Positron emission tomography (PET), and computed tomography (CT)	Enforces a desired trade-off between the false positives and negatives	Other ML algorithms could have been explored
4	[7]	Deep learning (DL)	IMDB	Improves by 56.38 % on the IMDB dataset and by 16.89 % and 34.76 % on the manusactured datasets in terms of the F1-score.	False positive and false negative rates are not computed.
5	[8]	SMOTE	Mammogram dataset	Performs better in terms of R_oC compared to undersampling only.	Not applied on other datasets.
6	[9]	Active learning	Synthetic data	Downsampling	No. of samples for downsampling is predefined.
7	[10]	Deep learning	Chest X-Ray data	F1-score is 93%	Hyperparameter tuning is random
8	[11]	Recognition-based approach, Cost-sensitive learning and boosting	Liver, Diabetes, Hepatitis, Pima	Range of F-measure is 67%-97%	Model does not perform better for some datasets
9	[12]	Cost-sensitive approach and a hybrid approach, Decision Tree, Support Vector Machine, Ensemble Methods	66 Datasets from Knowledge Extraction based on Evolutionary Learning (KEEL) repository	A_UC for most algorithms is upto 93%	Class overlapping problem is yet to be addressed
10	[13]	Modified AdaBoost	12 datasets from University of California Irvine (UCI) repository	Oversampling does not help AdaBoost	Indepth study of the parameters of AdaBoost
11	[14]	Review of data ugmentation techniques	Online Repository	GAN based augmentation is more powerful than handcrafted image-based techniques	Study is for only image dataset
12	[15]	Different forms of oversampling and deep learning	Modified National Institute of Standards and Technology database (MNIST), Canadian Institute for Advanced Research (CIFAR-10)	Effect of skewed data on classification model is detrimental	Only R_OC and A_UC metrics are focussed
13	[16]	Review on deep learning	Medical Image Data	Deep learning techniques are effective for medical imaging tasks	Limitations of methods are not focussed
14	[17]–[20]	Deep learning-based malware detectors	Corpus collected from VirusShare repositories	Accuracy of the model is about 93% for various malware data.	Model may be tsted for other type of malwares

Table 1. Summary table of related work (*continue*)

Sr#	Cite	Methodology	Dataset	Findings	Limitations
15	[21]	Transfer learning-based CNN-pretrained Visual Geometry Group (VGG-16), ResNet-50, and Inception-v3	Brain Tumor Cell	VGG-16 performs better compared to other models in terms of accuracy during training	Validation accuracy is 90%. It can be further improved
16	[22]	Review on ontology-based recommender system	Collected from literature	Context based recommender is universally used.	Recommender systems are not tested by their performances
17	[23], [24]	Active learning	Drug drug interaction (DDI), Hallmarks of cancers corpus (HoC), Chemprot	Active Learning outperforms passive learning for unlabelled data classification	Time complexity of the model is not described
18	[25]	Apriori algorithm	Lung cancer	Strong association of disease factors	Degree of association is not investigated
19	[26]	Temporal CNN	Seismic data	Cost sensitive CNN and ResNet outperform multilayer perceptron (MLP) and long short-term memory (LSTM) except precision metric.	Lack of multivariate benchmark dataset
20	[27]	Random Forest, Logistic Regression, Linear support vector classifier (SVC), Hybrid Sampling, SMOTE	Lung cancer	Sampling techniques improves classification accuracy. Oversampling is better than undersampling in terms of accuracy.	Accuracy is not the classification measure for imbalanced dataset
21	[28]	Deep learning	Mammogram image	DL performs better in terms of accuracy	Model is not tested on multiple datasets
22	[29], [30]	Review on CNN in radiology	Literature study	DL has powerful impact on medical imaging in addition to radiology	An incorporation of a small error can misclassify an input image.
23	[31]	Random forest regression, Automated hyperparameter optimization	Stroke	The false positive rate, accuracy and sensitivity predicted by the proposed approach are respectively 33.1, 71.6, and 67.4%.	There is further scope of improvement in false positive rate and sensitivity
24	[32]	Data Preparation	Oral squamous cell carcinoma	A histopathological image repository of normal epithelium of oral cavity and oral squamous cell carcinoma	Imbalanced Dataset
25	[33]	Apriori method, FP-Growth method	Mesothelioma disease	Prognostic Factors obtained through Association Rule Mining methods and validated by support, confidence and lift.	Application of proposed framework on a large dataset will be time expensive.
26	[34]	Classifiers like decision tree, AdaBoost etc. alongwith sampling techniques	Ovarian cancer	Decision tree using Support Vector Machine SMOTE (SVMSMOTE) has the most robust predictive ability.	Sensitivity of the model is less.
27	[35]	AlexNet, InceptionV3, and RegNetY-320 with data augmentation techniques	Skin Cancer	RegNetY-320 outperformed InceptionV3 and AlexNet in terms of the accuracy, F1-score.	Accuracy, F1-score, and ROC curve value obtained with the proposed framework were 91%, 88.1%, and 0.95.
28	Current study	Oversampling, CNN with two convolution layers of 16 and 64 kernels respectively, Pooling, ReLu and ANN for binary classification	OSCC (Mendley Dataset)	Accuracy is 99%, Precision score is 0.98, Recall score is 0.98, F1score is also 0.98, and Area under curve is 0.99.	The data used for the study is Mendley dataset. Proposed model may be tested for real time medical data for external validity.

REFERENCES

[1] Y. Qiang and W. Xindong, “10 Challenging problems in data mining research,” *Int. J. Inf. Technol. Decis. Mak.*, vol. 5, no. 4, pp. 597–604, Dec. 2006, doi: 10.1142/S0219622006002258.

[2] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.

[3] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[4] M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, “Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance,” *Neural Networks*, vol. 21, no. 2–3, pp. 427–436, Mar. 2008, doi: 10.1016/j.neunet.2007.12.031.

[5] S. Vluymans, “Learning from imbalanced data,” in *Studies in Computational Intelligence*, Springer International Publishing, 2019, pp. 81–110. doi: 10.1007/978-3-030-04663-7_4.

[6] S. A. Taghanaki *et al.*, “Combo loss: Handling input and output imbalance in multi-organ segmentation,” *Comput. Med. Imaging Graph.*, vol. 75, pp. 24–33, Jul. 2019, doi: 10.1016/j.compmedimag.2019.04.005.

[7] J. Jang, Y. Kim, K. Choi, and S. Suh, “Sequential targeting: A continual learning approach for data imbalance in text classification,” *Expert Syst. Appl.*, vol. 179, p. 115067, Oct. 2021, doi: 10.1016/j.eswa.2021.115067.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[9] W. Lee and K. Seo, “Downsampling for Binary Classification with a Highly Imbalanced Dataset Using Active Learning,” *Big Data Res.*, vol. 28, p. 100314, May 2022, doi: 10.1016/j.bdr.2022.100314.

[10] J. H. Chan and C. Li, “Learning from imbalanced COVID-19 chest X-ray (CXR) medical imaging data,” *Methods*, vol. 202, pp. 31–39, Jun. 2022, doi: 10.1016/j.ymeth.2021.06.002.

[11] G. Hoang, A. Bouzerdoum, and S. Lam, “Learning Pattern Classification Tasks with Imbalanced Data Sets,” in *Pattern Recognition*, InTech, 2009. doi: 10.5772/7544.

[12] V. López, A. Fernández, J. G. Moreno-Torres, and F. Herrera, “Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics,” *Expert Syst. Appl.*, vol. 39, no. 7, pp. 6585–6608, Jun. 2012, doi: 10.1016/j.eswa.2011.12.043.

[13] S. Wang and X. Yao, “Multiclass imbalance problems: Analysis and potential solutions,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 42, no. 4, pp. 1119–1130, Aug. 2012, doi: 10.1109/TSMCB.2012.2187280.

[14] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, vol. 6, no. 1, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[15] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” *Neural Networks*, vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.

[16] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, “Medical Image Analysis using Convolutional Neural Networks: A Review,” *J. Med. Syst.*, vol. 42, no. 11, Oct. 2018, doi: 10.1007/s10916-018-1088-1.

[17] K. Shaukat, S. Luo, and V. Varadharajan, “A novel method for improving the robustness of deep learning-based malware detectors against adversarial attacks,” *Eng. Appl. Artif. Intell.*, vol. 116, p. 105461, Nov. 2022, doi: 10.1016/j.engappai.2022.105461.

[18] K. Shaukat *et al.*, “Performance comparison and current challenges of using machine learning techniques in cybersecurity,” *Energies*, vol. 13, no. 10, p. 2509, May 2020, doi: 10.3390/en13102509.

[19] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, “A Survey on Machine Learning Techniques for Cyber Security in the Last Decade,” *IEEE Access*, vol. 8, pp. 222310–222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[20] K. Shaukat, S. Luo, S. Chen, and D. Liu, “Cyber Threat Detection Using Machine Learning Techniques: A Performance Evaluation Perspective,” in *1st Annual International Conference on Cyber Warfare and Security, ICCWS 2020 - Proceedings*, IEEE, Oct. 2020. doi: 10.1109/ICCWS48432.2020.9292388.

[21] C. Srinivas *et al.*, “Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images,” *J. Healthc. Eng.*, vol. 2022, pp. 1–17, Mar. 2022, doi: 10.1155/2022/3264367.

[22] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. M. Alam, and S. Luo, “A Review of Content-Based and Context-Based Recommendation Systems,” *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 3, pp. 274–306, Feb. 2021, doi: 10.3991/ijet.v16i03.18851.

[23] U. Naseem, M. Khushi, S. K. Khan, K. Shaukat, and M. A. Moni, “A comparative analysis of active learning for biomedical text mining,” *Appl. Syst. Innov.*, vol. 4, no. 1, p. 23, Mar. 2021, doi: 10.3390/asi4010023.

[24] S. Kamran *et al.*, “The Impact of Artificial intelligence and Robotics on the Future Employment Opportunities,” *Trends Comput. Sci. Inf. Technol.*, pp. 050–054, Sep. 2020, doi: 10.17352/tcsit.000022.

[25] M. Z. Latif, K. Shaukat, S. Luo, I. A. Hameed, F. Iqbal, and T. M. Alam, “Risk Factors Identification of Malignant Mesothelioma: A Data Mining Based Approach,” in *2nd International Conference on Electrical, Communication and Computer Engineering, ICECCE 2020*, IEEE, Jun. 2020. doi: 10.1109/ICECCE49384.2020.9179443.

[26] Y. Geng and X. Luo, “Cost-sensitive convolutional neural networks for imbalanced time series classification,” *Intell. Data Anal.*, vol. 23, no. 2, pp. 357–370, Apr. 2019, doi: 10.3233/IDA-183831.

[27] M. Khushi *et al.*, “A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data,” *IEEE Access*, vol. 9, pp. 109960–109975, 2021, doi: 10.1109/ACCESS.2021.3102399.

[28] S. Tripathy and R. Singh, “Convolutional Neural Network: An Overview and Application in Image Classification,” in *Advances in Intelligent Systems and Computing*, Springer Nature Singapore, 2022, pp. 145–153. doi: 10.1007/978-981-16-4538-9_15.

[29] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: 10.1007/s13244-018-0639-9.

[30] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[31] T. Liu, W. Fan, and C. Wu, “A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset,” *Artif. Intell. Med.*, vol. 101, p. 101723, Nov. 2019, doi: 10.1016/j.artmed.2019.101723.

[32] T. Y. RAHMAN, L. B. MAHANTA, C. CHAKRABORTY, A. K. DAS, and J. D. SARMA, “Textural pattern classification for oral squamous cell carcinoma,” *J. Microsc.*, vol. 269, no. 1, pp. 85–93, Jan. 2018, doi: 10.1111/jmi.12611.

[33] T. M. Alam *et al.*, “A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining,” *Biomed. Signal Process. Control*, vol. 68, p. 102726, Jul. 2021, doi: 10.1016/j.bspc.2021.102726.

[34] X. Yang, M. Khushi, and K. Shaukat, “Biomarker CA125 Feature Engineering and Class Imbalance Learning Improves Ovarian Cancer Prediction,” in *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2020*, IEEE, Dec. 2020. doi: 10.1109/CSDE50874.2020.9411607.
[35] T. M. Alam *et al.*, “An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset,” *Diagnostics*, vol. 12, no. 9, p. 2115, Aug. 2022, doi: 10.3390/diagnostics12092115.

BIOGRAPHIES OF AUTHORS

Usha Manasi Mohapatra holds a Doctor of Philosphy degree in Computer Science from S’O’A Deemed to be University, Odisha, India in 2019. She also received her B.Sc. (Mathematics) and MCA from Utkal University, Odisha, India in 1999 and 2005, respectively. She completed her MTech in Computer Science from BPUT, Odisha, India in 2011. She is currently an assistant professor at Computer Science Department in G.M. University, Odisha, India. She is having around 16 years of experience in teaching. Her research interest includes financial forecasting, ditributed learning, applied machine learning, and image classification. She has published papers in international journals and conferences. She can be contacted at email: ushamanasimohapatra@gmail.com.

Sushreeta Tripathy received the Ph.D. degree in Computer Science and Engineering from S’O’A Deemed to be University, Bhubaneswar, India. She is currently working as Assistant Professor in the Department of Computer Application, Faculty of Engineering and Technology Siksha ‘O’ Anusandhan Deemed to be University. At present, she has more than 16 years of teaching experience in the field of Computer Science and Engineering. Dr. Tripathy's principal research interest is Machine learning, Image processing, IoT, Big data health analytics and medical image analysis. Her aspiration is to work at the interface of these different fields or Multidisciplinary Environment. She is IEEE, IET and IAENG member. She has organized many conferences as well as workshops successfully. She can be contacted at email: sushreetatripathy@soa.ac.in.