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

Usha Manasi Mohapatra¹, Sushreeta Tripathy²
¹Department of Computer Science, Gangadhar Meher University, Sambalpur, Odisha, India
²Department 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.

This is an open access article under the CC BY-SA license.

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 imbalanced 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 (AUC) respectively. The main objective of our study is to minimize the failure rate in classification for class imbalance dataset. By inspective 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
An efficient convolutional neural network-based classifier for an imbalanced ... (Usha Manasi Mohapatra)
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 filters 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](image)

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 (R0C)-AUC, 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.

President 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-N. 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.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  

(1)

Table 2. List of acronyms used for classification performance metrics

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive (Type 1 Error)</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative (Type 2 Error)</td>
</tr>
<tr>
<td>TP R</td>
<td>True Positive Rate</td>
</tr>
<tr>
<td>FP R</td>
<td>False Positive Rate</td>
</tr>
<tr>
<td>TN R</td>
<td>True Negative Rate</td>
</tr>
<tr>
<td>FN R</td>
<td>False Negative Rate</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under Curve</td>
</tr>
</tbody>
</table>

An efficient convolutional neural network-based classifier for an imbalanced ... (Usha Manasi Mohapatra)
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.

$$\text{Precision} = \frac{T_P}{T_P + F_P}$$

(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 $F_NR$) is defined as the proportion of actual negatives, which got predicted as the negative.

$$\text{Recall} = \frac{T_P}{T_P + F_N}$$

(3)

$$\text{Specificity} = \frac{T_N}{T_N + F_P}$$

(4)

$$F_1\text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

(5)

$$F_PR = \frac{F_P}{T_P + F_P}$$

(6)

$$F_NR = \frac{F_N}{T_N + F_N}$$

(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

<table>
<thead>
<tr>
<th></th>
<th>Predicted Negative</th>
<th>Predicted Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Negative</td>
<td>True Negative ($T_N$)</td>
<td>False Positive ($F_P$)</td>
</tr>
<tr>
<td>Actual Positive</td>
<td>False Negative ($F_N$)</td>
<td>True Positive ($T_P$)</td>
</tr>
</tbody>
</table>

Another important metric is the ROC 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 $F_NR$ (shown in X-axis). In some literature, it is also suggested to take other parameters along X-axis. An example of ROC curves is shown in Figure 6.

Figure 6. Demonstration of ROC curves

It is a curve that is plotted taking different pairs of ($T_PR$, $F_NR$) values for different classification thresholds. In the Figure 3 different ROC curves are plotted. Each curve corresponds to one classifier. The area under a particular ROC curve is termed as $A_{ROC}$. It is marked with slanted lines in the Figure 6. The classifier
An efficient convolutional neural network-based classifier for an imbalanced... (Usha Manasi Mohapatra)

whose ROC curve covers a large area is considered as a better classifier. As per this convention, the classifier corresponding to top ROC 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 un-augmented datasets are provided in Figures 7(a)-7(b) and Figures 8(a)-8(b).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>No. of Normal Images</th>
<th>No. of Carcinoma Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>89</td>
<td>439</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>201</td>
<td>495</td>
</tr>
</tbody>
</table>

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 imbalance, the classification performance is poor in terms of FPR. The poor classification performance can be observed from the ROC plot which is depicted in Figure 9(a). The confusion matrix representing the FP, TP, FN, and TN 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 ROC plot drawn for the augmented dataset 1 using FPR and TPR of the proposed classifier. The confusion matrix generated for the classifier is depicted in Figure 10(b).

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

Figure 8. Loss Plot during training for Un-augmented data; (a) dataset 1 and (b) dataset 2
Figure 9. Performance measures for un-augmented dataset 1; (a) ROC and (b) confusion matrix

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

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

It can be seen that the AUC for the ROC 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 ROC and confusion matrix are portrayed in Figures 11(a) and 11(b) respectively. ROC 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, FPR, and FNR comparison for both unaugmented and augmented data considered in dataset 1. Similarly, Figure 13(b) illustrates Accuracy, FPR, and FNR 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 FPR and FNR. In the healthcare decision-making process, false predictions are very hazardous.
An efficient convolutional neural network-based classifier for an imbalanced dataset (Usha Manasi Mohapatra)

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

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

**Figure 13. Bar plot demonstrating accuracy, FPR and FNR; (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 contingency table. In this article the contingency 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 worthy 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 defined 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.
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

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

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


BIOGRAPHIES OF AUTHORS

Usha Manasi Mohapatra holds a Doctor of Philosophy 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, distributed learning, applied machine learning, and image classification. She has published papers in international journals and conferences. She can be contacted at email: ushamanasi.mohapatra@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: sushreeta.tripathy@soa.ac.in.