Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 26 (2023), No. 1, pp. 67–86
DOI : 10.47974/JSMS-948

A predictive model for bankruptcy: ANN, LSTM and CNN approaches

Sasmita Manjari Nayak †
Minakhi Rout *
School of Computer Engineering
Kalinga Institute of Industrial Technology (Deemed to be University)
Bhubaneswar
Odisha
India

Abstract

Predicting bankruptcy is the focus of our research, which is one of the important aspects of research, as due to bankruptcy both company's goodwill and shareholders' benefits are affected. In order to predict bankruptcy, reliable models are required. The focus of this paper is based on different deep learning models. However, developing deep learning models for forecasting bankruptcy is one of the challenging tasks as most of the datasets are imbalanced in nature. So we first try to balance the dataset. US Bankruptcy Prediction Data set (1971-2017) is taken here, which is very imbalanced in nature. To balance the dataset both undersampling and oversampling and one hybrid method are used. In this research, a comparison is made among three different types of models like CNN, LSTM and ANN by applying all balancing techniques on each classifying model, for the prediction of bankruptcy. Here we get that the ANN model gives better results than the other two whatever balancing technique may be used and among the balancing techniques oversampling is far better than undersampling, but here the hybrid sampling method outperformed all in each classification model as compared to the others.

Subject Classification: *68: Computer Science.*

Keywords: *Bankruptcy, Undersampling, Oversampling, SMOTE, SMOTETOmek, Artificial neural networks(ANN), Recurrent neural network(RNN), Long short term memory networks (LSTM), Convolutional neural network(CNN).*

† *E-mail:* sasmita.lina@yahoo.com
* *E-mail:* minakhi.routfcs@kiit.ac.in (Corresponding Author)

1. Introduction

Bankruptcy is a legal procedure that allows companies to get rid of their debts when a business is unable to meet its financial commitments [1]. The term "bankruptcy prediction" refers to the process of predicting bankruptcy using a number of financial factors [2]. Failure of a company can never come overnight so forecasting provides a warning before time to take appropriate actions. For this region bankruptcy prediction has been the subject of much academic and professional research over the past half-century.

To train the bankruptcy model huge amounts of historical data are needed. In most cases, the dataset used is imbalanced in nature. The unequal distributions among different classes in a dataset, it is called imbalance. A prediction model built on an unbalanced dataset may be ignored by the minority group. This issue may influence the efficiency of the models that have been developed. Several methods are there to handle the imbalanced issue. Overall all these balancing techniques are of two types: undersampling and oversampling. The goal of undersampling is to eliminate the data from the dataset whereas in oversampling, fresh samples are generated in the minority class instances until the majority class and the minority class is having the same number of data.

In our study, we take the US Bankruptcy Prediction Data set (1971-2017) to simplify bankruptcy decision-making. It is a highly imbalanced dataset, so first, we try to balance it by applying one undersampling method and two oversampling methods. We make a comparison among the different sampling methods. We also tried a hybrid sampling method which was constructed by the combination of oversampling method SMORT and an undersampling method. Finally, we construct bankruptcy models for forecasting financial distress. We concentrate on three different types of classification methods in this study to forecast bankruptcy. We construct CNN (Convolutional Neural Network), LSTM (Long Short Term Memory) and ANN (Artificial Neural Network) models and compare them by applying each balancing technique on each model. Here the confusion matrix, the accuracy score, the f1_score, the precision_score and the recall_score are used as the performance measure.

2. Literature Review

Forecasting of bankruptcy helps finance sectors not to finance the companies that may go bankrupt in near future; also the investors can

turn away financial loss by avoiding the investment in such companies. The prediction rate depends on both data imbalance issues and prediction methods. Bankruptcy prediction methods are of two types, one is statistical methods and another is artificial intelligence methods. The first attempt for Prediction of bankruptcy through statistical methods by doing different financial ratio analyses was done in 1966. Later Muti Dimensional Analysis(MDA) was used for this purpose [3,4]. Statistical methods have some limitations like they have low uncertainty tolerance, cannot operate on very large datasets, etc. To overcome such types of limitations many researchers go for different artificial intelligence methods. In general, the artificial intelligence method includes two groups, machine learning methods and Deep Learning Methods (DL) [5],[6]. In [7],[8] by comparing machine learning models and traditional models, it is found that models based on machine learning have been shown to be extremely effective in finance applications, as compared to traditional methods. In [9],[10] by comparing among different machine learning models it is found that one of the most effective machine learning models was the Support Vector Machine (SVM). In [11] it is found that the forecasting power of machine learning techniques like Neural Networks (NN) is based on the presence of a significant amount of data to train which means a large dataset is needed. For which it is also a difficult task to get an appropriate NN model. To overcome the above limitations some researchers try to apply deep learning for bankruptcy prediction. In [12],[13] to avoid explicit extraction of features and to develop implicitly learning by the learning algorithms deep learning is applied for corporate default prediction and compare them with machine learning methods. Here it is found that the deep learning techniques are better than traditional machine learning techniques. Deep learning models can also handle very large datasets. A Deep Dense Multilayer Perceptron model can predict bankruptcy accurately. Whereas the authors show successful results in image classification by applying CNN [14-17]. In [18], after making a comparison among two different types of deep learning methods (average embedding and CNN) and three different types of other machine learning models on basis of textual data, it is found that deep learning algorithms outperform traditional models in predicting bankruptcy based on textual data also. The prediction capability of RNN and LSTM methodologies can outperform all traditional machine learning models [19,20]. Researchers compared logistic regression and ANNs models and found out that the ANN gives better results than the logistic regression when a larger dataset is used for testing [21],[22].

In [2],[23] it is found that SVM was less sensitive to imbalanced datasets. When balancing data with SMOTE, the outcomes are better than random undersampling for machine learning models. Oversampling is a better choice as compared to undersampling, as it is ideal for all kinds of forecasting models. In [24] it is found that the sampling method depends on the size and type of data in the dataset. For a few numbers of bankrupt cases, SMOTE is better, but if there is present huge number of bankruptcy cases, undersampling is preferable to oversampling.

Some hybrid methods are also introduced by some researchers to balance the imbalanced dataset. In [25] Borderline SMOTE is combined with Stacked AutoEncoder (SAE), which gives better results than machine learning models for bankruptcy prediction. The hybrid method named SMOTETomek is applied by some authors to balance the dataset [26],[1]. In [27] a hybrid model named GA- ANN is introduced by combining the Genetic Algorithms (GA) and ANN.GA-ANN outperformed ANN with random sampling while using cluster-based evolutionary under-sampling.

In this literature review, we got different types of classification methods. Here we may conclude that Artificial Intelligence (AI) techniques are somewhat preferable to statistical techniques. Among different types of AI methods, deep learning techniques are preferable to machine learning techniques. But we did not get any work for finding the best method for bankruptcy prediction. For which we try to make a comparison among CNN, LSTM and ANN and try to find out the best one for bankruptcy prediction by taking only the numerical data into consideration. Through this literature review, we also got different types of sampling techniques. We also got some comparisons among them but no one tried to get the most suitable sampling method for deep learning models and ANN. In this research, we will try to find out the most appropriate sampling technique for balancing a highly imbalanced dataset for deep learning models and ANN by comparing some simple balancing techniques. Here we consider one undersampling method i.e. random undersampling, two oversampling methods i.e. random oversampling, SMOTE and one hybrid method SMOTETomek.

3. Methodology

Our methodology is having different stages like Dataset Description, Data Preprocessing, Classification Models, etc.

3.1 *Dataset Description*

The data we utilize comes from the US Bankruptcy Prediction Data set (1971 to 2017). It has 15 attributes like Liquidity, Profitability, Productivity, Leverage Ratio, Asset Turnover, Operational Margin, Return on Equity, Market Book Ratio, Assets Growth, Sales Growth, Employee Growth, etc and 92872 records. One can get this dataset by using this link https://www.kaggle.com/shuvamjoy34/us-bankruptcy-prediction-data-set-19712017.

3.2 *Data Preprocessing*

Data pre-processing can be a very complex and time-consuming phase and the predictive results are very much dependent on the type of procedure followed. Preprocessing of data can be said as a technique through which the data of a dataset can be given a more usable form. There are several dataset preprocessing techniques, which are so helpful to make better prediction accuracy. Balancing to the imbalanced dataset, handling the outlier mainly comes under data preprocessing. In this paper, we only consider the imbalanced nature of the dataset.

Practically as compared to non-bankrupt cases, the bankrupt cases are extremely less. That's why a severe imbalance problem arises. So data balancing is necessary before classification. Through the review we get that, mainly there are present two principal types of balancing techniques one is oversampling and another is undersampling. Both oversampling and undersampling are called resampling. In this paper following resampling techniques are considered.

I. Random Under-Sampling (RU)

The undersampled dataset is nothing but just a part of the original dataset which is created by deleting some data from the majority class. In RU data from the majority class is randomly eliminated till the sample size of both minority class and majority class becomes equal. In random undersampling, the size of the dataset decreases and the useful data may be eliminated from the original dataset. This may result in poor performance if the size of the original dataset is small in itself [23].

II. Random Oversampling (RO)

In random oversampling, some instances are chosen from the minority class, then replicated them again and again after that they are

added into the same minority class. This procedure is continued until both the classes are having an equal number of instances. Mainly this technique is used by the researchers for balancing the dataset as there is no loss of data occurring like undersampling. Random oversampling increases the size of the dataset. No loss of data but, an overfitting problem may occur, as here duplicate copies of the same minority class data are created [23].

III. Synthetic-Minority-Over-sampling-Technique (SMOTE)

SMOTE artificially produces fresh data for the minority class not using replicated data as random oversampling to balance the dataset. So in SMOTE, the overfitting problem is avoided. In SMOTE, a member of the minority class is chosen. The k-nearest neighbors are then found out by using a distance measurement approach. It inserts newly created fresh data somewhere along the line for each minority data with any of its nearby minority-class neighbors [23].

IV. Smote + Tomek links

In oversampling we can balance minority class and majority class, but skewed class distributions problems still remain in datasets. Though in SMOTE synthesized or artificially created data are added in the minority class to balance the dataset, still overfitting problems may occur. For which, Tomek links can be used as a data cleaning method on the over-sampled training set. In Tomek links, data from both minority and majority classes are removed which form Tomek links [28].

3.3 *Classification Models*

In this paper, we focus on three types of classification models: Artificial Neural Networks (ANN), Convolution Neural Networks (CNN), Long Short-Term Memory (LSTM). Here each of them is discussed in more detail.

I. Artificial Neural Networks (ANN)

The ANN models are very flexible and non-linear in nature. ANN is a feed-forward neural network. Both the structure and function of ANN are the replication of the human brain. In general, ANN is having three types of layers i,e. an input layer, one or more hidden layers and an output layer. In ANN output depends on the input as well as on bias value. The hidden neuron's output can be calculated by applying the following formula:

$$o_h = f^{(i)}(b_i^{(i)} + \sum_{j=1}^{in} wih_{ij} x_i) \quad (1)$$

Where x_i is the input data, i =1, ..., n; and wih_{ij} denotes the weight connecting input neuron x_i to hidden neuron j. The output of the output neurons is calculated as follows:

$$o_o = f^{(ii)}(b_i^{(ii)} + \sum_{j=1}^{hn} wo_{ij} h_j) \quad (2)$$

Where hn is the number of hidden neurons and wo_{ij} is the weight associated with hidden neuron h_j. The networks are initially set up using random weights, then the weights are gradually modified to minimize the loss function [25].

II. Long Short Term Memory Networks (LSTMs)

The LSTM is having a better architecture to decide whether the information is relevant or not. That means the information can be remembered or irrelevant one can be forgotten could be decided by LSTM. Mainly the LSTM consists of three parts. The first part decides the relevancy of the information coming from the previous timestamp. If relevant that will be remembered or if irrelevant that should be forgotten. That's why this part is known as forget gate. The second part is called the input gate and the third part, is called the output gate.The equation for the forget gate.

$$F_t = \sigma(i_t * w_t + h_{t-1} * W_h) \quad (5)$$

Where,

σ = sigmoid function

i_t = input to the current timestamp.

w_t = weight associated with the input

h_{t-1}= The hidden state of the previous timestamp

W_h = It is the weight matrix associated with the hidden state

The value of F_t lies between 0 and 1 . This value is multiplied by the previous timestamp's cell state.

$c_{t-1} * F_t = 0$, if $F_t = 0$ (Here data is irrelevant, so forget everything)

$c_{t-1} * F_t = c_{t-1}$, if F_t =1 (Here data is relevant, so remember everything)

If F_t is 0 then the network will remember nothing and if the value of F_t is 1 it will remember the information.

The input gate is being used to measure the significance of fresh data carried by the input. Mathematically it can be represented as follows:

$$I_t = \sigma(i_t * w_i + h_{t-1} * W_i) \quad (6)$$

Where,

i_t = Input at the current timestamp t

w_i = weight matrix of input

h_{t-1} = A hidden state at the previous timestamp

W_i = Weight matrix of input associated with hidden state

After applying the sigmoid function over it, the value of I_t that is at timestamp t, the value of input which lies between 0 and 1. If I_t is 0 then the network will remember nothing and if the value of I_t is 1 it will remember the information. Now it is the time to pass on the fresh information to the cell state which is a function of a hidden state at timestamp t-1 and input I at timestamp t [31].

III. Convolutional Neural Network (CNN)

CNN may have a number of convolution layers, as it is having convolution layers the network is very deep with fewer parameters. CNN is also known as ConvNets. In CNN input is of fixed length whereas outputs are independent of previous inputs. The spatial properties of an image are taken into consideration by CNN. The arrangement of pixels in an image, as well as their connection, are referred to as spatial characteristics. This aids in the correct identification of the object accurately.To apply the CNN it is essential to modify the data structure into a 3D matrix. Each unit should correspond to a data matrix [32].

There is no need of applying filtering methods explicitly as CNN learns the filters automatically. These filters can extract the suitable characteristics from the data. A Rectified Linear Unit (ReLU) layer is also present in CNNs, which is used to conduct operations on components. As a result, a feature map is created. that has been rectified. Then the rectified feature map sent into a pooling layer through which the dimensions of the feature map is reduced. Through this process by flattening the pooled feature map, it was turned into a single, long, continuous, linear vector. Finally,when the pooling layer's flattened matrix is presented as an input, a completely linked layer forms, which categories and identifies the images.

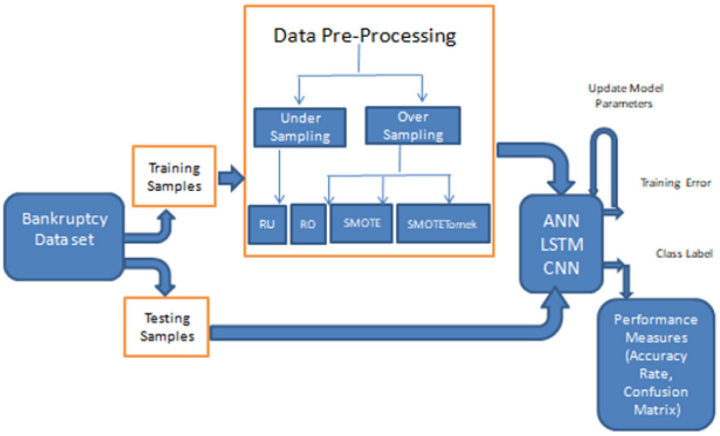

Figure 1

The architecture of our bankruptcy prediction process.

3.4 *Architecture of Proposed Model*

The process of construction of our bankruptcy prediction models and their evaluation can be represented pictorially step by step as shown in Figure 1.

4. Results

First the dataset is splited into two parts: training and testing. The datasets that were used in this study were split: 90% for training the models and 10% to evaluate the efficiency of the models. It is very imbalanced in nature as in training set out of 90000 data only 158 bankrupt cases are present. To balance the dataset, different types of balancing techniques are used here. Both undersampling and oversampling methods are applied and make a comparison among them. In undersampling, only random undersampling is used whereas in oversampling both random oversampling and SMORT are applied and simultaneously one hybrid method SMOTETOmek is used for balancing the imbalanced training dataset where, a data cleaning method, Tomek linkages are applied to the SMOTE over-sampled training set. After that, the classification models are created. Then the balanced training set gets into the classification models to train the model. After the final trained model is created,the performance is evaluated by testing it against the test set.We use classification performance

metrics such as confusion matrix, ROC curve, accuracy_score, f1_score, precision_score and recall_score to evaluate the efficiency of the model.

Here we are using three types of classification models for bankruptcy prediction i,e. ANN,CNN and LSTM. All four sampling methods I,e.random undersampling, random oversampling, SMOTE and SMOTETOmek are applied to each model and make comparisons among them to find out which sampling method is best and which classification model is best among these three models.

In the first part of our experiment, we should set up the classification model parameters asthese significantly influence the outcome of the model. The most important variables or parameters that influence the performance are the max epoch, the number of hidden layers, learning rate, batch size, the number of neurons in the hidden layer, etc. For our implementation, we are using Python 3.7(TensorFlow) platform.

Our ANN model is havinga single input layer, a single hidden layer and a single output layer. There are 15 neurons in the input layer as the dataset has 15 feature columns. The output layer has 2 neurons as we need two outputs i.e. bankrupt and nonbankrupt. We are taking only one hidden layer which is having 32 neurons. To compile the model 'sparse_categorical_crossentropy' type of loss function is used. In order to optimize the loss,an Adam optimizer is used with a learning rate of 0.0001. We set up the batch size to 10. and the number of epochs to100 for training the model.

From the confusion matrix shown below, it is clearly found out that, for the testing dataset which is having 2472 nonbankrupt cases the ANN model with SMOTETOmek balancing technique is able to identify all in all but out of 400 bankrupt cases it is able to identify only 384 cases correctly, with 16 errors. Whereas the ANN model with the Random undersamplingbalancing technique gives thelowestresult. From the confusion matrix shown below, it is found out that, for the testing dataset out of 2472 nonbankrupt cases 819 cases are identified and out of 400 bankrupt cases only 294 cases can be identified correctly. Among the Random oversamplingand SMORT sampling methods, by applying SMORT, ANN gives better results. After using SMOTE all non-bankrupt cases are identified whereas 379 bankrupt cases are identified correctly. But when using the random oversampling method all bankruptcy cases are identified correctly but only 372 bankrupt cases areidentified correctly. Figure 2 and Figure 3 shows the confusion matrices and ROC curveobtained from ANN by applying different sampling methods respectively.

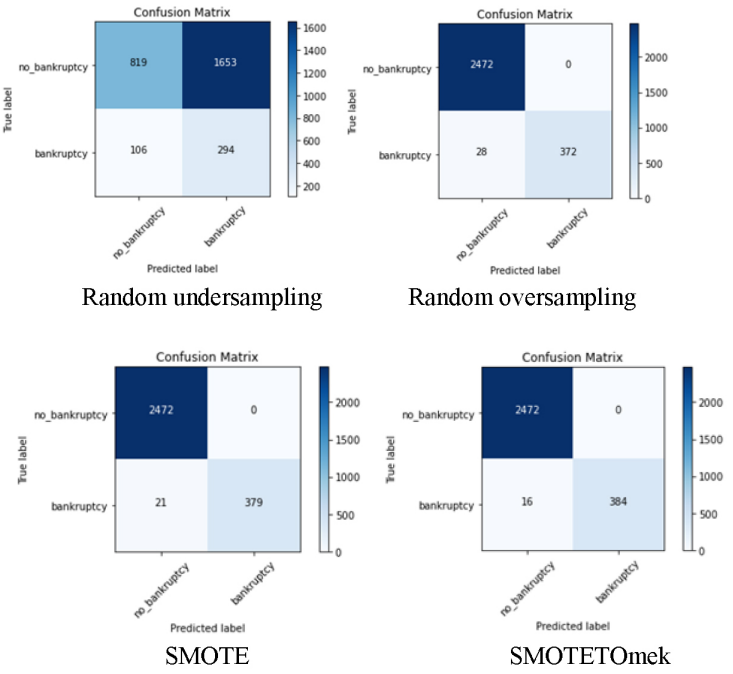

Figure 2

Confusion Matrix obtained from ANN for Dataset 1.

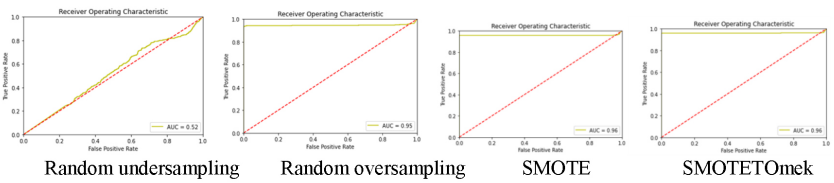

Figure 3

ROC Curve obtained from ANN for Dataset 1.

Our LSTM model is a sequentially dense model. The first LSTM layer has 100 neurons. We are using the input shape as our training set's form. One more layer was then added. Finally the output layer. We start with the LSTM layer, then a few Dropout layers are added, to avoid overfitting. We specify 0.2 for the Dropout layers, which means 20% of the layers will be removed.The Dense layer is next added, which specifies a single output unit. Then Adam optimizer is used to compile the model and the sparse

categorical_cross_ entropy is used as the loss function. The model was then trained using the fit function having 100 epochs whose batch size is 32.

From the confusion matrix shown below, it is clearly found out that, for the testing dataset with SMOTETOmek balancing technique, out of 2472 nonbankrupt cases the LSTM model is capable to detect 2245 cases correctly, with 227 errors and out of 400 bankrupt cases the LSTM model is capable to detect all most all cases correctly, with only 4 errors. Whereas the LSTM model with the Random undersampling balancing technique

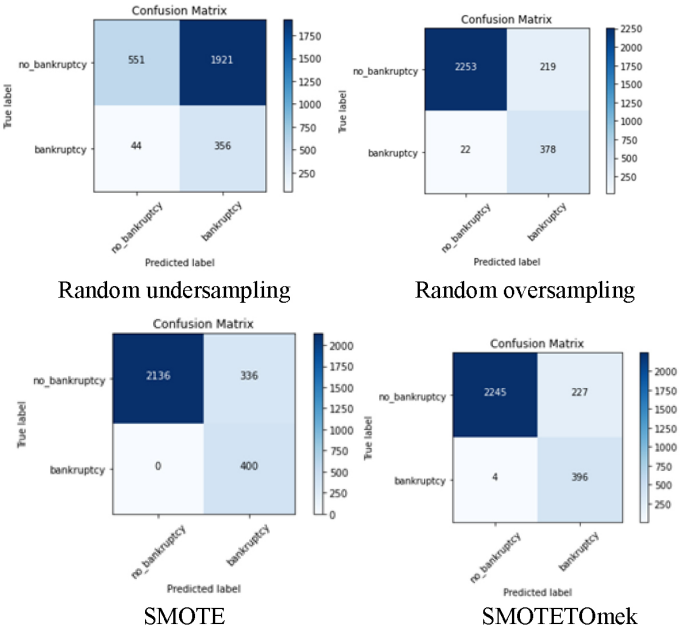

Figure 4

Confusion Matrix obtained from LSTM for Dataset 1.

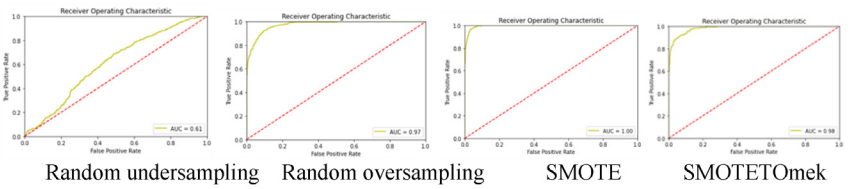

Figure 5

ROC Curve obtained from LSTM for Dataset 1.

gives the lowest result. From the confusion matrix, it is found out that, out of 2472 nonbankrupt cases 551 are identified and out of 400 bankrupt cases only 356 cases can be identified correctly. Among the Random oversampling and SMORT sampling methods, applying SMORT, LSTM gives better results. After using SMORT 2136 non-bankrupt cases are identified whereasallbankrupt cases are identified correctly.But when using the random oversampling method 378 bankruptcy cases are identified correctly but only 2253 non-bankrupt cases are not identified correctly. Figure 4 and 5 shows the confusion matrices and ROC curves obtained from LSTM by applying different sampling methods respectively.

The CNN model has Conv1D(). So it is having a one-dimensional convolution layer. In the first Conv1D() layer, over the dataset,we apply

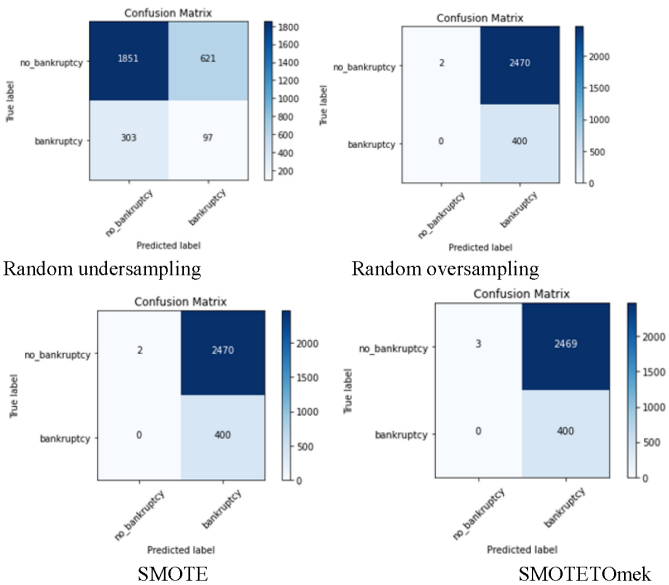

Figure 6

Confusion Matrix obtained from CNN for Dataset 1.

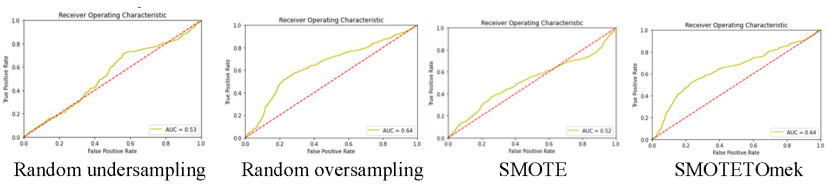

Figure 7

ROC curve obtained from CNN for Dataset 1.

32 filters, with the convolutional window having a size of 3. The input_ shape parameter specifies the input's shape. Finally, four layers of hidden neurons with rectified linear activation function (ReLU) activation function are present. After that, the output comes, which is a dense layer with 2 neurons because our predicting value is either 0 or 1. Here the softmax function is employed because it ranges from 0 to1, making it easier to forecast a binary outcome. In this model Flatten() is used to convert the data into a one-dimensional array for further processing to the next layer. To compile and fit the model,an Adam optimizer with a 0.0001 learning rate is used. 100 epochs are taken to train the model.

Like the other two classifying models here the CNN model with SMOTETOmek balancing technique also gives thebest resultas compare the other balancing techniques. From the confusion matrix shown below, it is clearly found out that, for the testing dataset with SMOTETOmek balancing technique, out of 2472 nonbankrupt cases the CNN model is able to identify only 3 cases correctly, but all bankrupt cases are identify correctly, without any error. Whereasthe CNN model with the Random undersampling balancing technique gives the lowest result. Here out of 2472 nonbankrupt cases 1851 are identified but out of 400 bankrupt cases only 97 cases can be identified correctly. Among the Random oversampling and SMOTE sampling methods, by applying SMORT, CNN gives better results. After using SMORT only 2 non-bankrupt cases are identified whereas out of 400, all bankrupt cases are identified correctly. Figure 6 and Figure 7 shows the confusion matrices and ROC curves obtained from CNN by applying different sampling methods respectively.

Table 1

Accuracy-score of different models on different balancing techniques for Dataset 1.

Predictive Models	Synthesis data generation techniques			
	Random undersampling	Random oversampling	SMOTE	SMOTETOmek
ANN	0.3875	0.9902	0.9926	0.9944
LSTM	0.3158	0.9160	0.8830	0.9195
CNN	0.6782	0.1399	0.1399	0.1403

Table 2

F1_score of different models on different balancing techniques for Dataset 1.

Predictive Models	Synthesis data generation techniques			
	Random undersampling	Random oversampling	SMOTE	SMOTETOmek
ANN	0.3663	0.9790	0.9844	0.9881
LSTM	0.3126	0.8537	0.8156	0.8626
CNN	0.4868	0.1231	0.1231	0.1235

Table 3

Precision_score of different models on different balancing techniques for Dataset 1.

Predictive Models	Synthesis data generation techniques			
	Random undersampling	Random oversampling	SMOTE	SMOTETOmek
ANN	0.558	0.9944	0.9957	0.9967
LSTM	0.5411	0.8117	0.7717	0.8169
CNN	0.5109	0.5759	0.5696	0.5697

Table 4

Recall_score of different models on different balancing techniques for Dataset 1

Predictive Models	Synthesis data generation techniques			
	Random undersampling	Random oversampling	SMOTE	SMOTETOmek
ANN	0.5920	0.9650	0.9737	0.98
LSTM	0.5564	0.9282	0.9320	0.9490
CNN	0.4956	0.5004	0.5004	0.5006

From the above resultant Tables 1 - 4, we can easily compare our three predictive models and four sampling methods. In each predictive model, the accuracy rate, the f1_score, the precision_score and the recall_ score are high for the hybrid preprocessing method SMOTETOmek and simultaneously the confusion matrixshows that the number of correct predictive cases of both non-bankrupt cases as well as bankruptcases is

also high. For the random undersampling method, the accuracy rate, the f1_score, the precision_score and the recall_score are least in each and every predictive model. Here also from the confusion matrix, it is clearly found out the number of correct predictive cases (TP i,e. true positive and TN i,e. true negative) cases are also least for each and every predictive model. Among the other two oversampling methods i,e. SMORT and random oversampling, SMOTE always gives better results than random oversampling based on accuracy rate, the f1_score, the precision_score and the recall_score as well as correct predictivecases (TP and TN). So here it can be concluded that oversampling methods are better than undersampling methods. Among the three oversampling methods discussed here i,e. the random oversampling method,SMOTE and SMOTETOmek, thehybrid method SMOTETOmekis the best one for balancing the imbalanced dataset. So here we may conclude that for balancing the imbalanced dataset oversampling is better than undersampling and among the oversampling methods the hybrid method SMOTETomek is the best one. The undersampling method is not suitable for any method as the undersampling method always gives the worst result.

Among the three predictive models,for each sampling method, ANN shows the highest accuracy score, the f1_score, the precision_score and the recall_score. CNN shows the highest accuracy score and f1_score only for the random undersampling technique but it's precision_score and the recall_score are the lowest as compared to other classification models. So by observing the result of the above tables here we may conclude that the ANN model with SMOTETOmek sampling method is the best for bankruptcy prediction. The confusion matrix for the testing dataset alsoshows that the ANN appears to be very successful in predicting the bankruptcy as compared to other models with the SMOTETOmek sampling method, as out of 2472 nonbankrupt cases the ANN model is able to identify all in all and out of 400 bankrupt cases the ANN model is able to identify 384 cases correctly, with only 16 errors. ANN shows a high accuracy_score, the f1_score, the precision_score and the recall_ score,which is much more than the other two models.So it is very clear that the model ANN is more appropriate for bankruptcy prediction as compared to CNN and LSTM, whereas the CNN model is not preferable as compared to the other two models for bankruptcy prediction.

5. Conclusion

The primary goal of our research was to anticipate bankruptcy by using deep learning algorithms. We are using the US Bankruptcy dataset which is highly imbalanced for which common classifiers cannot give proper prediction results by processing such datasets. In this paper, we observe the influence of one under-sampling method and two oversampling methods and one hybrid method to deal with the imbalanced datasets. Our findings show that, in general, over-sampling techniques, always gives better prediction result as compared to the undersampling method, but among oversampling methods SMORT gives better results than the random oversampling method and finally, we found that the hybrid method Smote + Tomek gives the best results in practice for each discussed method.

Three different models. (ANN, CNN and LSTM) are used here. The ANN gives a high accuracy_score, f1_score, precision_score and the recall_score and outperformed other models on each sampling technique, whereas CNN shows the worst predictive result.

In this paper, we only consider the imbalanced nature of the dataset and try to balance the dataset. We did not consider the outlier data of the dataset. So in the future, we should work on the outlier data and will try to remove them by applying some required techniques. In the future, we may consider some more types of deep learning methods and will try to find out the best one for the prediction of bankruptcy in deep learning.

References

[1] Vellamcheti, S., & Singh, P. (2020, January). Class Imbalance Deep Learning for Bankruptcy Prediction. In 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T) (pp. 421-425). IEEE.

[2] Alam, T. M., Shaukat, K., Mushtaq, M., Ali, Y., Khushi, M., Luo, S., & Wahab, A. (2020). Corporate bankruptcy prediction: An approach towards better corporate world. *The Computer Journal.*

[3] Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research,* 71-111.

[4] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.

[5] Chen, M. Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications,* 38(9), 11261-11272.

[6] Yu, L., Yang, Z., & Tang, L. (2016). A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment. *Flexible Services and Manufacturing Journal,* 28(4), 576-592.

[7] Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417

[8] Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science,* 162, 895-899.

[9] Shin, K. S. Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications,* 28(1), 127-135.

[10] Erdogan, B. E. (2013). Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. *Journal of Statistical Computation and Simulation,* 83(8), 1543-1555.

[11] Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1), 127-135.

[12] Yeh, S. H., Wang, C. J., & Tsai, M. F. (2014, July). Corporate default prediction via deep learning. In Proceedings of the 34th International Symposium on Forecasting (ISF'14), Rotterdam, The Netherlands (Vol. 29).

[13] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data,* 2(1), 1-21.

[14] Alexandropoulos, S. A. N., Aridas, C. K., Kotsiantis, S. B., & Vrahatis, M. N. (2019, May). A deep dense neural network for bankSruptcy prediction. In International Conference on Engineering Applications of Neural Networks (pp. 435-444). Springer, Cham.

[15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.

[16] Bengio, Y., & LeCun, Y. (2007). Scaling learning algorithms towards AI. *Large-scale Kernel Machines*, 34(5), 1-41.

[17] Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications,* 117, 287-299.

[18] Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743-758.

[19] Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems,* 104, 38-48.

[20] Kim, H., Cho, H., & Ryu, D. (2021). Corporate Bankruptcy Prediction Using Machine Learning Methodologies with a Focus on Sequential Data. *Computational Economics*, 1-19.

[21] Youn, H., & Gu, Z. (2010). Predict US restaurant firm failures: The artificial neural network model versus logistic regression model. *Tourism and Hospitality Research*, 10(3), 171-187.

[22] Park, S. S., & Hancer, M. (2012). A comparative study of logit and artificial neural networks in predicting bankruptcy in the hospitality industry. *Tourism Economics,* 18(2), 311-338.

[23] Veganzones, D., & Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, 112, 111-124.

[24] Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. Knowledge-Based Systems, 41, 16-25.

[25] Smiti, S., & Soui, M. (2020). Bankruptcy prediction using deep learning approach based on borderline SMOTE. *Information Systems Frontiers,* 22(5), 1067-1083.

[26] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synt hetic minority over-sampling technique. *Journal of Artificial Intelligence Research,* 16, 321-357.

[27] Kim, H. J., Jo, N. O., & Shin, K. S. (2016). Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. Expert Systems with Applications, 59, 226-234.

[28] Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD explorations newsletter, 6(1), 20-29.

[29] Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research*, 222(1), 168-178.