Mapana - Journal of Sciences
2023, Vol. 22, Special Issue 2, 223-244
ISSN 0975-3303|https://doi.org/10.12723/mjs.sp2.12

Imbalanced Multiclass Data Classification Using Combined Data Sampling and Deep Learning Method

S.Sivasankari*, Dr.S.Sukumaran*

Abstract

Multiclass Classification for finding pattern refers to classifying each data to part of the classes or labels that are generally more than two. The general strategies so far followed in Multiclass are reconstructing Multiclass to several binary classes with state of art methods such as One-versus-Rest, One-versus -One, Error Correcting Output codes, etc. The foremost challenge in classifying is with imbalanced data. It may exist in Binary as well as in Multiclass, but has a huge impact in the later method. This type of data has skewed portions of classes that have large portion known to be majority class, and small portion known as minority class. In this situation, the classifier would have more samples from majority class and not much from minority class that leads to poor understanding of samples and less accurate results. The existing works discussed Random Upsampling, Random

* Ph.D Research Scholar, 2Associate Professor,

* 2Erode Arts and Science College, Erode, Tamilnadu, India 1sivamirtha@gmail.com

Downsampling, SMOTE methods individually with FeedForward Neural Network and found Random Oversampling gave better results than other methods .However, it generates more duplicate data and has less accuracy. Hence to overwhelm these issues, this research work put forward Combined Random Over-Under Sampling approach in training data that was preprocessed prior with Replacing Missing value with mean, Feature selection and Noise Filtering. Meanwhile this work extends the existing FeedForward Neural Network to Deep Learning (Deep FeedForward Neural Network). The proposed work is implemented in Rapidminer tool, assessed with appropriate evaluation measures for training and testing data individually.

Keywords:-Imbalanced Data, Deep Learning, Stratified Sampling, Optimization Algorithm, Combined Random Over-Undersampling.

Introduction

Generally, in Data Mining while classifying the data, two types of target variables namely binary class, and multi-class are dealt with. [1] The Binary target variable has two categories while the Multiclass target variable has more than two categories in it. In today's scenario, there exists more number of categories or classes present in the target variable in most of the applications. The Logistic regression, support Vector Machine do not support for Multi-Class Classification. But SVM algorithm is extended now to support for it. The classifiers such as Decision tree classification, KNN, Naive bayes, Neural network-based models can handle this type of Classification however some fails to give superior performance. [2]. Even though some classifiers support directly or use the extended methods to handle it, there exists a common issue with Imbalanced data.

1.1 Method for Multiclass Classification

The multiple binary classification models are named as one-vs-one, one-vs-rest, and Error Correcting output codes.

One-vs-Rest- The method applies one binary model for every target(class) against all other targets.

One-vs-One-The method applies one binary model for each pair of targets.

Error Correcting Output codes- When the number of categories in the classes is smaller than seven, then exhaustive codes with the length 2^{k-1}-1 can be applied or else random codes are used.

1.2 Imbalanced Classification

This type of classification refers [3] to the phenomena where the number of samples in each class is distributed unevenly. The class with more number of samples refers to majority and with less number of samples is called minority class. This leads to unequal distribution in the training set that have less number of training samples for minority that in turn gives poor learning model and less predictive performance. Now-a-days, many real applications have this uneven distribution of data such as detecting fraud and spam, predicting churn etc. The distribution can vary from a slight bias to a severe. [4]The degree of imbalance may vary mild to extreme as shown in Table 1 with the proportion of minority class over the dataset.

Table 1. Degree of Imbalance with Proportion of Data

Degree of Imbalance	Proportion of Minority class over total data
Mild	20-40%
Moderate	1-20%
Extreme/Severe	<1%

Annotation by Authors

"An imbalance [5] exists when one or more classes/categories have very less portion among other categories."

"Class imbalance [6] is defined based on a particular dataset and it is typically gauged with respect to the training distribution."

"Any dataset with an [7] uneven class frequency is imbalanced. This is because of the disproportion among the samples of all targets."

"Developments in learning [8] from imbalanced data is inspired by numerous real applications. In such cases the minority usually require methods to improve its rates."

Imbalance level in between Classes - The Eq.(1) indicate the maximum inbetween –class imbalance in [18] where maxi(C_i) implies the maximum class size (majority class among the whole classes in the dataset) and $min_i(C_i)$ returns the minimum class size of overall class. Generally above the value 10 is treated as minority class.

$$\textbf{Imbalance IB} = \max_i(C_i)/\min_i(C_i) \quad (1)$$

Causes - The are two main reasons are the way the samples were collected and the features or nature of the domain itself.

Challenge – A mild imbalance is acceptable but the moderate and severe one can be the challenging and may require the use of improved methods.

Approaches

There are three methods to overcome this type of data such as (a) Data-level methods that modifies the collection of examples to balance the distributions using Over or undersampling, (b) Algorithm-level methods that modifies the existing learning algorithms to handle skewed distributions, (c) Hybrid methods that combine the advantages of both prior level methods.

2 Literature Review

Haseeb Ali *et al* [9] did a comprehensive survey in handling imbalanced class problems. In addition, the issues that endorse bias for majority, minority class was discussed. The work discussed about four types of approach namely Preprocessing approach, Algorithmic approach, Cost sensitivity approach and Ensemble learning to handle classification.

Farhan Ullah *et al* [10] proposed advance Loss function (Huber Loss) in deep neural network with collaborative filtering method to handle data sparsity, inaccuracy problem for education service recommendation. Four hidden layers (256, 128, 64, 32) neurons with Adam optimization algorithm was used on goodbook dataset taken from Kaggle repository. The results showed that Deep neural network with L2 regularization, ReLU activation function gave 0.60 MAE value that is least error from other methods.

Mustafa Bogal *et al* [11] presented milk yield classification model using deep neural network with cross validation. Data were collected from 156 Holstein cattle. Trinomial classification with good, poor, medium was taken for analysis. 7 hidden layers with neurons 64-12-256-512-256-128-64 respectively was fixed. Sigmoid activation, cross entropy loss function were applied and from the results it was shown that 76.92% accuracy was obtained for the *K* value 2.

Chittem Leela Krishna *et al* [12] dealt with Heart Disease data. Decision trees, NB, SVM and Deep Neural network was applied for the classification. The results proved that neural network composed with 3 hidden layers with neurons (15, 12, 5), sigmoid, MSE functions obtained 85.47% accuracy.

Máximo Eduardo Sánchez-Gutiérrez *et al* [13] proposed a machine learning neural network model with sigmoid function for healthcare data with two components namely restricted Boltzmann machine and a classifier system. It uses a discriminant pruning method. The results showed the error rate is minimized significantly by 7%.

Waleed *et al* [14] paper dealt with the problems such as imbalanced-overlapping datasets that often encounter synchronously. Three popular classification algorithms such as Decision Tree, KNN, and SVM were used for analysis. Two kinds of data sets balanced and unbalanced data sets with different overlapping and separation levels were taken. It was noted that PCA analysis appeared to be a good measure for the degree of balancing the datasets and all the three algorithms have very similar performance indicators.

ShujuanWang *et al* [15] put forth an improved SMOTE based on Normal distribution to avoid the marginalization. The healthcare big data are often taken for evaluation. The results showed improved method gave 2% more AUC value for all the datasets.

Johnson *et al* [16] did a survey that highlights various gaps in deep learning. The work discussed about the categories to handle data namely Data level method with Oversampling, Undersampling, Algorithm level method namely weight based algorithms, cost sensitive based algorithms and finally hybrid method with the combination of Data level, Algorithm level methods. The evaluation metrics for classification is also listed.

Dhanalakshmi *et al* [17] improved the predictive accuracy of classification in ordinal dataset. Data Structure based Oversampling was proposed in Training data set. The experimental analysis was performed on testing data set of variant benchmark datasets and highest accuracy 90.66% was obtained for bondrate dataset.

Justin M. Johnson *et al* [18] examined existing deep learning techniques. This survey discussed about the implementation and experimental results for the study. The survey concluded, highlighted various research gaps in deep learning that guide for future research.

3 Methodology

3.1 Existing Methods

The classification method [19] introduced an efficient prediction model MLP_ADAM using Multi-Layer Perceptron enhanced with training algorithm Adam, Tangent activation, Mean squared Error loss functions. The model was evaluated on three public educational datasets gathered from repositories. Synthetic Minority Oversampling Technique and random oversampling was assessed and concluded Random oversampling gave better results.

The method[20] implemented Weight Guided Wrapper Feature Subset method PWGWFS-DL using Deep Feed Forward Neural Network in order to reduce Curse of Dimensionality, minimize Computation Complexity, and to increase the accuracy on Data Classification. After Feature selection method 10 attributes out of 20 on sales transaction dataset was selected. Four hidden layers with 75 neurons in each layer was applied for Deep Learning. For handling imbalanced data, the work used Random Upsampling/Oversampling method.

Research Gaps in Existing Methods

The existing work used Data level method for classification. The existing methods [19] [20] suggested Random Oversampling that duplicates the data in minority class alone to handle data. However, multiple duplicate data in the minority class overfit (low bias and high variance) the model.

3.2 Proposed method

Preprocessing - The proposed method initially applies preprocessing method namely Replace Missing value with mean value on independent and dependent variables. Weight Guided Wrapper Feature Subset Method [20] is used to select features. Split Validation is used to split the data with the ratio (90%, 10%) for training and testing. The data are sampled using stratified sampling that ensures even class distribution.

Edited Nearest Neighbor (ENN)- This method is used to remove noisy data and so this is considered as noise filter. The noise data cannot be understood and interpreted correctly by the classifier. It is a meaningless data as compared with other instances. The method starts by finding the K-nearest neighbor (K) of every instance. Then check whether the majority class from the instance's k-nearest neighbor is unique as the particular instance's class or not, if it differs then the instance and its nearest neighbors (based on k value) are deleted from the dataset. To find nearest neighbor, any type of distance measure is used but most often Euclidean distance is used.

Combined Random Over-Undersampling with Deep Learning

The proposed method concept is to apply a Random Oversampling method to the minority class at the same time applying Random Undersampling on the majority class that reduce the bias and overfitting problem. In addition, it avoids adding more data for the complete classification.

Deep Learning with Multilayer Feed Forward Neural Network with Combined Random Data Sampling method for Imbalanced data (CDS-DLM)

Training set - The existing works FeedForward neural network is improved using combined Over-Undersampling, Deep Feedforward Neural Network using AdaDelta optimizing algorithm and analyzed for variant hidden layers with neurons, and dropouts.

Random Over-Undersampling methods are combined to make even distribution of the classes in training data. The Input layer neurons based on the independent variables, and the output layer based on the dependent variable or outcomes. Optionally dropouts, which mean dropping some ratio of neurons in hidden layer while processing that is not usually done in simple neural network. This work analyzes variant hidden layer structure and its neurons with dropouts. The epoch is set as 100. An improvement to traditional gradient descent algorithms the advanced method adaptive gradient descent optimization algorithms is utilized to avoid tuning of learning rates for optimal convergence. Epsilon and rho that are similar to learning rate and momentum are set as 1.0E-8, 0.99 respectively for Adaptive gradient descent. The Weight is (random initialization value 0.01) applied to the inputs along with the bias (value 1) while the inputs are transmitted between neurons using Rectifier Linear Unit (ReLU) Activation Function. The Huber Loss Function is used to compare the target and predicted output values errors. L2/Ridge Regularization is added as a penalty term. Finally, the errors are passed using Back Propagation method to train feed forward neural networks with early stopping. Dropouts implies dropping out the nodes in a neural network usually done in the input and hidden layers. But, this work applies only in hidden layer.

Testing set – The testing is implemented with Bagging ensemble meta algorithm with trained Deep Feed Forward Neural Network utilized in training set. The Bootstrap aggregating helps to enhance the performance by dealing with bias-variance, overfitting.

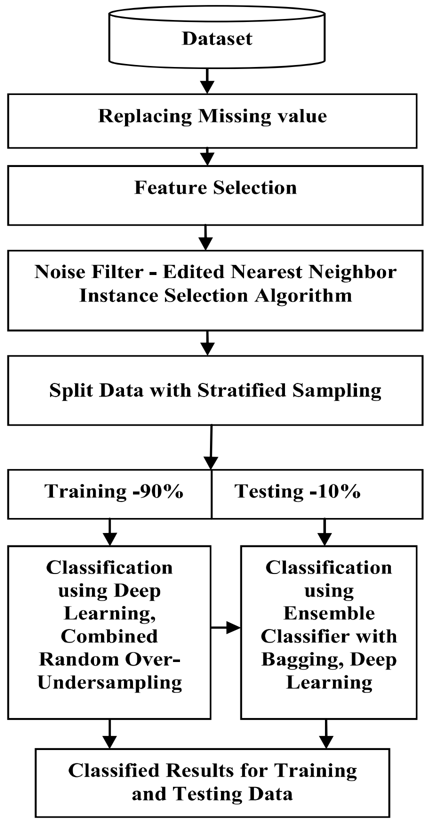

Fig 1. Flow of Work

Figure 1, denotes the steps of proposed work CDS-DLM. The dataset is preprocessed with replacing missing value, and relevant features are selected using PWGWFS-DL [20]. Edited Nearest Neighbor instance selection algorithm with class weights is applied to filter the noisy data. The whole dataset is split into training and testing with the ratio 90%, 10 % respectively using stratified sampling method. Then, the proposed combined Random Over-Undersampling method based Deep Learning is applied to the training data. This results are then used for testing data. The testing data is applied with Ensemble learner Bagging and trained Deep Learning method combined along random over-under sampling. Finally, the performance of both training and testing data are assessed with appropriate evaluation measure.

Algorithm of CDS-DLM Method

Input	*Imbalanced Multiclass Dataset*
Output	*Classified data*
Step 1:	*Replacing Missing Value with Mean, Feature Selection using PWGWFS-DL*
Step 2:	*Apply Edited Nearest Neighbor algorithm with K value 3* *For n number of instances,* *Select K-Nearest Neighbor for the instance and return the majority class.* *If the class of the instance and the majority of its neighbors vary, then the instance,its neighbors are deleted.* *End For* *Repeat the process of selecting neighbors and deleting the differed instances till the desired proportion of each class is completed.*
Step 3:	*Apply Split data method with Stratified Sampling in the proportion 90% for training and 10% for testing*
Step 4:	***Training Phase with Classifier*** *Train the model with Deep Learning Multi-Layer Feed Forward with the user parameter as follows,* *Add ratio to increase and decrease data on each class for Combined Random Over-Undersampling respectively.* *Initialization of Input, Hidden, Output layers, Weight (random initialization value 0.01) bias (value 1).* *The model produces the output o_j for each neuron j, with weights and bias as,* $o_j = Activation\ function\ (net_j)$ $For\ the\ input\ value\ j,$ $ReLU\ Activation\ Function =$ $f(x) = \begin{cases} 0, if\ j < 0 \\ x, if\ j \geq 0 \end{cases}$ $net_j = \sum_{k=1}^{n} w_{kj} x_k$ *The output is generated and Huber Loss function is used to determine the error E as,* $L_\delta(E) = \begin{cases} \frac{1}{2} E^2, & for \lvert E \rvert \leq \delta \\ \delta. \left(\lvert E \rvert - \frac{1}{2}\delta \right), & otherwise \end{cases}$

	L2/Ridge Regularization is added as a penalty term (value 0.01) by, $$L2\ regularization = L_\delta(E) + \lambda \sum_{i=1}^{N} w_i^2$$ *Calculate the partial derivatives with weight w_{ij} as,* $$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \cdot \frac{\partial o_l}{\partial net_l} \cdot \frac{\partial net_l}{\partial o_j} = \frac{\partial E}{\partial o_j} \frac{\partial o_l}{\partial net_l} o_i$$ $$\frac{\partial E}{\partial w_{ij}} = o_i \delta_j$$ *Update the weight w_{ij} using stochastic gradient descent with Adaptive Learning Rate as,* $$w_{ij} = w_{ij} - \eta \cdot \nabla_{w_{ij}} J(w_{ij};\ x^{\iota}; y^{\iota})$$ *Where Adaptive Learning rate υ_t is calculated as,* $$\upsilon_t = \gamma \upsilon_{t-1} + \eta \nabla_{w_{ij}} J(w_{ij}) \{0 < \gamma < 1\}$$ $$w_{ij} = w_{ij} - \upsilon_t$$ *Where J is the Huber loss function*
Step 5	***Testing Phase with Ensemble Classifier*** *Implement Ensemble meta-algorithm using Bagging as follows,* *Create m new training sets from D_i from the training sets D with replacement.* *Train the model with Deep Learning Multi-Layer Feed Forward with the user parameter initialized in training phase.* *The class with maximum votes is chosen as the predicted label.*
Step 6:	*Assess the performance of Training, Testing data individually.*

Advantages

The proposed method overcomes the issue in handling imbalanced data by combining Random Over- Undersampling in training data. This method avoids duplication of numerous data in the whole dataset in preprocessing. Test data accuracy is improved with the removal of Noise data using Edited Nearest Neighbor method, Stratified sampling, Ensemble classifier. The proposed method avoids bias and overfitting which is a regular issue if the methods Oversampling/Undersampling were applied in isolation.

4 Results and Discussion

4.1 Dataset Description

The Superstore Sales Order (SSO) Transaction dataset is taken from Kaggle repository. It consists of 20 attributes, one target attribute based on order priority namely High, Medium, Low, and Critical and 9988 instances.

Distribution of samples over the class

Table 2, shows the proportion of the four classes namely Medium, High, Critical and Low Among them, the class namely Critical and Low is in Moderate as well as in minority class.

Table 2. Distribution of Data

Class	Absolute Count	Fraction of data (Absolute Count/ Total Data
Medium	5561	0.557 – 55.7%
High	3118	0.312 – 31.2%
Critical	847	0.085 – 8.5%
Low	462	0.046 – 4.6%
Class Imbalance ratio as per Eq.(1)	12.0	

4.2 Preprocessing

The proposed method applies Weight Guided Wrapper Feature Subset method to get relevant feature set and obtained ten attributes. The combined Random Over-Undersampling ratio for the training data is fixed as 0.90, 0.95, 1.5, 2.0 for the classes High, Medium, Critical and Low classes respectively. After Noise filtering the dataset has 7440 data. Among them, the training set has 6696 data and test set has 744 data.

4.3 Performance Evaluation

The performance is evaluated using the following measures from Eq. (2) –Eq. (6). Accuracy, WMR, WMP are in percentage and RMSE, Kappa Statistics is in range between 0-1.

Accuracy

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{n} \quad (2)$$

Weighted Mean Recall (WMR)

$$\text{WMR} = \frac{1}{n}\sum_{i=1}^{n} C_i * \frac{TP_i}{TP_i + FN_i} \quad (3)$$

Weighted Mean Precision (WMP)

$$\text{WMP} = \frac{1}{n}\sum_{i=1}^{n} C_i * \frac{TP_i}{TP_i + FP_i} \quad (4)$$

Where, TP_i is the True-Positive value, TN_i is True-Negative value, FP_i is False-Positive value, FN_i is the False-Negative in Class C_i, n is total samples. C_i are the labels Medium, High, critical, Low in the SSO dataset.

Root Mean Square Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (5)$$

Where, n is total samples, $\hat{y}_i$ is predicted value, y is the observed value of i^{th} observation.

Kappa Statistics

$$\text{Kappa Statistic} = \frac{(P_o - P_E)}{1 - P_E} \quad (6)$$

Where, P_o is the proportion of observed agreement, P_E is the proportion of expected agreements by chance.

4.4 Results

1. Training Phase

In Table 3, the existing methods namely MLP_ADAM, PWGWFS-DL is compared with the proposed method CDS-DLM for accuracy. H1, H2, H3 H4 refers to the hidden layer size with 1,2 3, 4 and 50 neurons each. In Table 3, it is observed that the method CDS-DLM obtains higher results with 98.9%

for Hidden layer size three.

TABLE 3. Comparison of Methods for Accuracy on Training Data with variant Hidden Layer sizes

Method	Hidden layer Size			
	H1	H2	H3	H4
MLP_ADAM	89.8	90.8	92.0	90.1
PWGWFS-DL	92.6	93.9	95.1	93.2
CDS-DLM	94.5	96.2	98.9	95.5

2. Testing Phase

In Table 4, the existing and proposed methods is compared for accuracy and observed that the proposed method CDS-DLM obtains better results with 92.5 % for Hidden layer size three.

TABLE 4. Comparison of Methods for Accuracy on Testing Data with variant Hidden Layer sizes

Method	Hidden layer Size			
	H1	H2	H3	H4
MLP_ADAM	80.0	81.9	84.2	81.0
PWGWFS-DL	84.7	85.2	88.4	84.1
CDS-DLM	87.1	90.0	92.5	90.2

3. Analysis on using variant neurons with Hidden Layer size 3 for Training and Testing set

In Table 5, the proposed method is tried for variant number of neurons (25, 50, 75, 100) distributed uniformly on each layer. It confirms that accuracy is less for 25, 75, 100 neurons and hidden layer with 3, 50 neurons provides better results.

TABLE 5. CDS-DLM method for Accuracy (3 hidden layers with different number of neurons)

Data	Number of Neurons in Hidden layer			
	25	50	75	100
Training	90.9	98.9	93.4	92.0
Testing	79.8	92.5	84.3	82.0

4. Analysis on using variant Dropout ratios with Hidden layer size 3 for Training and Testing set

The proposed method is assessed for variant dropout ratio in the hidden layer in the following Table 6. Five variant ratios' are applied 0.10, 0.15, 0.25, 0.50 and 0.75. The resultant values in Table 6, proves that high dropout ratio leads to poor accuracy. The least dropout ratio 0.10 provides high accuracy but when compared to the same architecture in Table 3, 4 the accuracy level is slightly less for Training and Testing set. Therefore, it is concluded dropout ratio in neural network might be optional, if the network is built with optimal structure.

TABLE 6. CDS-DLM method with variant Dropout ratio on Training, Testing Data (3 hidden layers, 50 neurons)

Data	Dropout Ratio in Hidden layer				
	0.10	0.15	0.25	0.50	0.75
Training	97.6	96.2	95.1	75.4	55.6
Testing	89.9	89.0	87.4	69.0	55.6

5. Analysis on Existing and Proposed Methods with variant evaluation measures

Training Phase

In Table 7, the proposed method outperforms the existing by having high values 93.4%, 96.0%, 0.17, and 0.98 for WMR, WMP, RMSE and Kappa Statistics respectively.

TABLE 7. MLP_ADAM, PWGWFS-DL, CDS-DLM methods on Training data (3 hidden layer, 50 neurons each)

Method	WMR	WMP	RMSE	Kappa
MLP_ADAM	88.2	90.1	0.32	0.87
PWGWFS-DL	90.7	92.4	0.20	0.92
CDS-DLM	93.4	96.0	0.17	0.91

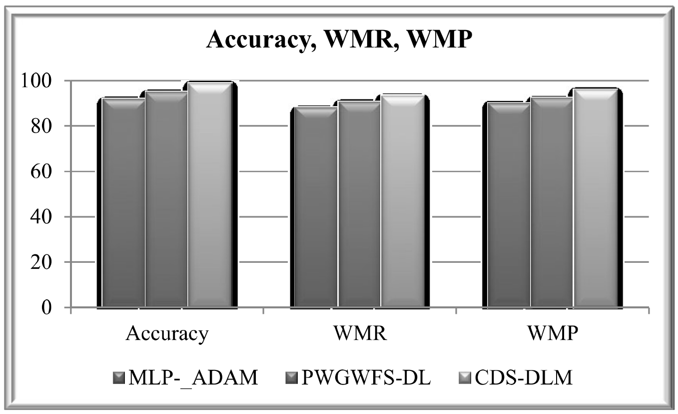

Fig 2. Accuracy, WMR, WMP for Training Data

In Figure 2, CDS-DLM obtains high accuracy, WMR and WMP on training data than PWGWFS-DL and MLP_ADAM methods.

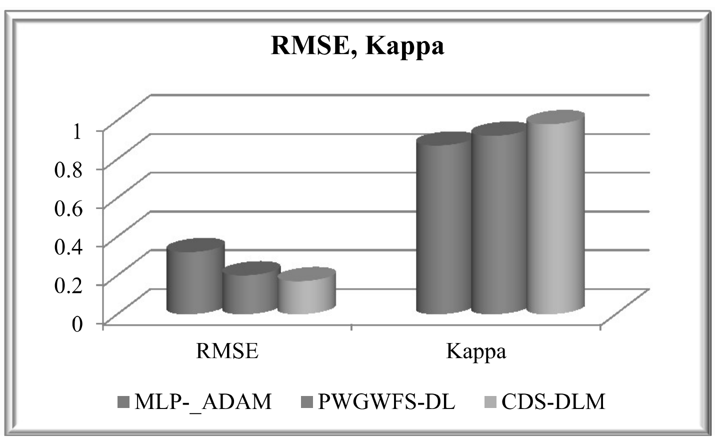

Fig 3. RMSE, Kappa for Training Data

In Figure 3, CDS-DLM has less RMSE and high Kappa values on training data that implies the proposed method efficiency.

Testing Phase

In Table 8, the proposed method outperforms the existing by having high values 89.1 %, 91.5%, 0.3 and 0.90 for WMR, WMP, RMSE and Kappa Statistics respectively.

TABLE 8. MLP_ADAM, PWGWFS-DL, CDS-DLM methods on Testing data (3 hidden layer, 50 neurons each)

Method	WMR	WMP	RMSE	Kappa
MLP_ADAM	80.7	83.2	0.51	0.78
PWGWFS-DL	85.1	86.9	0.47	0.84
CDS-DLM	89.1	91.5	0.30	0.86

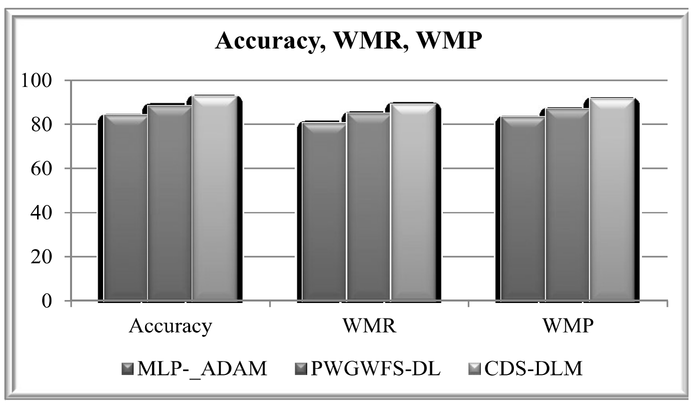

Fig. 4 Accuracy, WMR, WMP for Training Data

In Figure 4, CDS-DLM obtains high accuracy, WMR and WMP on testing data than PWGWFS-DL and MLP_ADAM methods.

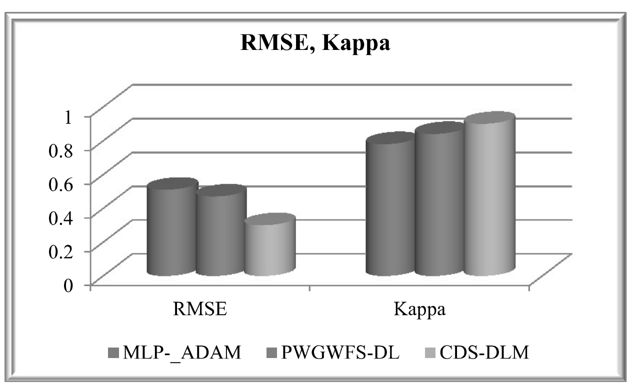

Fig 5. RMSE, Kappa for Testing Data

In Figure 5, CDS-DLM has less RMSE and high Kappa values on testing data that implies the efficiency of the proposed method.

5 Conclusion

Multiclass Classification on Imbalanced dataset often arises in many practical applications. It leads to less accurate results with existing classifiers. Mostly, the existing work applied Data-Level method namely Random Oversampling, Undersampling separately for this classification. With the aim to improve the performance in data classification, this work applied combined Random Over-Under Data sampling method. For classifying the data, Deep Learning was used but in this network fixing the size of hidden layers and its neurons is a challenging task. This work experimented hidden layers, its neurons, dropout ratio. The analysis is done on Transaction data with 90% training data and 10% testing data. The performance of the method CDS-DLM confirms that the hidden layer with size 3, 50 neurons obtains highest accuracy of 98.9%, 92.5% on training data and testing data respectively. Hence, the proposed work minimizes Curse of Dimensionality by using efficient feature subset algorithm, reduces the Computation Complexity with actual data, and increases the accuracy on Imbalanced Data in both training and test set and avoids overfitting.

In Future, this work can be applied for the applications such as Students' Academic performance, Healthcare.

References

1. Daniel T. Larose , Chantal D. Larose, "Data Mining and Predictive Analytics", Second Edition, Wiley Publishing, (2015).
2. Hussain Ahmad Madni; Zahid Anwar; Munam Ali Shah, "Data mining techniques and applications A decade review", 23rd International Conference on Automation and

Computing (ICAC), IEEE publication, (2017).

3. JafarTanha, YousefAbdi, NeginSamadi, NazilaRazzaghi, Mohammad Asadpour, “Boosting methods for multi-class imbalanced data classifcation: an experimental review”, Journal of Big Data, pp.1-47, (2020).
4. Pradeep Kumar, RoheetBhatnagar, Kuntal Gaur, AnuragBhatnagar, “Classification of Imbalanced Data:Review of Methods and Applications”, IOP Publishing, pp.1-9, (2021).
5. Max Kuhn, Kjeli Johnson, “Applied Predictive Modeling”, Springer, First edition, pp.-419, (2018).
6. Yunqian Ma, Haibo He, “Imbalanced Learning: Foundations, Algorithms, and Applications”, Wiley, pp.- 16, (2013).
7. Fernandez, “Learning from Imbalanced Datasets”, Springer, First Edition, pp -19, (2018).
8. Bartoszkrawczyk, “Learning from imbalanced data: open challenges and future directions”, Springer, pp-221-232, (2016).
9. Haseeb Ali, MohdNajibMohdSalleh, Rohmat Saedudin3, KashifHussain, Muhammad FaheemMushtaq, “Imbalance class problems in data mining: a review”, Indonesian Journal of Electrical Engineering and Computer Science Vol. 14, No. 3, pp. 1560-1571, (2019).
10. FarhanUllah, Bofeng Zhang, RehanUllah Khan, Tae-Sun Chung, Muhammad Attique, Khalil Khan, Salim El Khediri, Sadeeq Jan, “Deep Edu: A Deep Neural Collaborative Filtering for Educational Services Recommendation”, IEEE Access, volume 8, pp.110915-110928, (2020).
11. Mustafa Bogal , KerimKürşatÇevik, AykutBurgut, “Classifying Milk Yield Using Deep Neural Network”, Pakistan J. Zool., vol. 52(4), pp 1319-1325, (2020).
12. ChittemLeela Krishna Dr. PoliVenkataSubbaRedd, “An Efficient Deep Neural Network Multilayer Perceptron

Based Classifier in Healthcare System", 3rd International Conference on Computing and Communication Technologies ICCCT 2019, IEEE, pp.1-6, (2019).

13. Máximo Eduardo Sánchez-Gutiérrez, Pedro Pablo González-Pérez, "Multi-Class Classification of Medical Data Based on Neural Network Pruning and Information-Entropy Measures", Entropy, MDPI, pp.1-20, (2022).
14. Waleed A. Almutairi and RyszardJanicki, "On relationships between imbalance and overlapping of datasets", Proceedings of 35th International Conference on Computers and Their Applications, EPiC Series in Computing Volume 69, pp. 141–150, (2020).
15. ShujuanWang, Yuntao Dai, JihongShen&JingxueXuan, "Research on expansion and classifcation of imbalanced data based on SMOTE algorithm", Scientific Reports, (2021).
16. Johnson, J.M., Khoshgoftaar, T.M. Survey on deep learning with class imbalance. Journal of Big Data 6, 27, (2019).
17. Dhanalakshmi .D , Anna SaroVijendran, "Adaptive Data Structure Based Oversampling Algorithm for Ordinal Classification", Indonesian Journal of Electrical Engineering and Computer Science Vol. 12, No. 3, , pp. 1063~1070, (2018).
18. Justin M. Johnson, Taghi M. Khoshgoftaar, "Survey on deep learning with class imbalance", Journal of Bigdata, (2019).
19. ThaerThaher, Rashid Jayousi, "Prediction of Student's Academic Performance using Feedforward Neural Network Augmented with Stochastic Trainers", 14th International Conference on Application of Information and Communication Technologies, IEEE, pp.1-7, (2020).
20. Sivasankari .S, Dr.S.Sukumaran, Dr.S.Muthumarilakshmi, "Deep Learning based Weight Guided Wrapper Feature Subset Method for Multiclass Data Classification", NeuroQuantology, Volume 20, Issue 16, , pp.5675-5686, (2022).