Preprint arXiv:1901.06091v1 [cs.LG]

Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry

Uzair Ahmed[1], Asifullah Khan[1], Saddam Hussain Khan[1], Abdul Basit[1,2], Irfan Ul Haq[1], and Yeon Soo Lee[3]*

[1]Pattern Recognition Lab, Department of Computer & Information Sciences, Pakistan Institute of Engineering & Applied Sciences, Nilore, Islamabad 45650, Pakistan

[2]TPD, Pakistan Institute of Nuclear Science and Technology, Nilore, Islamabad, Pakistan

[3]Department of Biomedical Engineering, College of Medical Science, Catholic University of Daegu, South Korea

asif@pieas.edu.pk

Abstract: A churn prediction system guides telecom service providers to reduce revenue loss. Development of a churn prediction system for a telecom industry is a challenging task, mainly due to size of the data, high dimensional features, and imbalanced distribution of the data. In this paper, we focus on a novel solution to the inherent problems of churn prediction, using the concept of Transfer Learning (TL) and Ensemble-based Meta-Classification. The proposed method “TL-DeepE” is applied in two stages. The first stage employs TL by fine-tuning multiple pre-trained Deep Convolution Neural Networks (CNNs). Telecom datasets are in vector form, which is converted into 2D images because Deep CNNs have high learning capacity on images. In the second stage, predictions from these Deep CNNs are appended to the original feature vector and thus are used to build a final feature vector for the high-level Genetic Programming and AdaBoost based ensemble classifier. Thus, the experiments are conducted using various CNNs as base classifiers with the contribution of high-level GP-AdaBoost ensemble classifier, and the results achieved are as an average of the outcomes. By using 10-fold cross-validation, the performance of the proposed TL-DeepE system is compared with existing techniques, for two standard telecommunication datasets; Orange and Cell2cell. In experimental result, the prediction accuracy for Orange and Cell2cell datasets were as 75.4% and 68.2% and a score of the area under the curve as 0.83 and 0.74, respectively.

Keywords: Churn Prediction, Deep Convolutional Neural Networks, Telecom, Transfer Learning, Meta-Classification, Genetic Programming, Ada-Boost, Imbalance Data, High Dimensionality, Customer Retention.

1. Introduction

‘Customer churn’ is a term related to the customer subscription model and refers to the process, whereby a customer terminates to use the services of a provider [1]. This term is frequently used in the telecommunication industries [2, 3] since several factors like coverage area and lucrative offers from competitors can trigger a user’s decision to switch from one service provider to another. Telecom corporations can benefit greatly by retaining customers that are on the verge of churning because it has been estimated that the cost of retention is usually 5 to 15 times less as compared with the cost of acquiring a new customer [4, 5]. The corporations, therefore, need an effective churn prediction system to automate the task of detecting the customers who might churn. The prediction is usually made on customer’s usage patterns and other factors, which directly or indirectly affect the user’s decisions.

Customer Churn Prediction is a complex problem having challenges such as messy data, Low churn rate, Churn event censorship, etc. [6] The current customer churn prediction methods are largely based on machine learning based classification methods. However, the performance of a classifier is generally suffered due to high dimensions of the telecommunication datasets. Furthermore, the telecommunication data has an imbalanced nature in its distribution (with a limited number of examples of the minority class) that also hampers in achieving accurate results. Churner prediction is considered as a binary classification problem, where churners lie in the minority class and non-churners lie in the majority class. Several machine learning algorithms focus on improving the whole classification performance by sacrificing the accurate prediction of churners in the minority classes. Generally, churners (minority class samples) are more misclassified as compared with non-churners, due to the fundamental problem of its imbalanced nature. Due to this reason, machine learning based classification has a tendency to classify all samples as non-churners, which can provide good accuracy but less precision in terms of predicting churner’s class. The complication shows that there is still improvement by a large margin, keeping in view of the new challenges and opportunities that are arising in the domain of big data analytics.

Previously, Support Vector Machines (SVMs) were applied to improve the performance of telecommunication churn prediction. Similarly, Filter and Wrapper methods are collectively used for feature selection, which is further provided to ensemble classifiers. These ensemble classifiers include Random Forest, RotBoost and SVMs [7, 8] that explore the selected feature vectors helpful for churn classification as reported by Idris et al.[9]. Yabas et al. presented metaclassification methods [10] for standard telecom Orange dataset. Their selection of meta-classification methods includes bagging, Random Forest, Logistic Regression and Decision Trees, which reported an Area Under the Curve (AUC) of 0.7230. Along similar lines, Koen et al. applied Rotation Forest and AdaBoost-based ensemble and reported an AUC as 0.697[11]. Verbeke et al. have proposed different methods including SVMs, Naïve Bayes, k-NN, Neural Network and reported AUC as 0.714 on standard telecommunication Orange dataset [12]. Similarly, a ensemble classifier with mRMR based processing [13] reported that RotBoost and Random Forest-based ensemble[14] resulted in an AUC of 0.749. The approach of GP and AdaBoost ensemble [15] was presented for churn prediction telecommunication data and reported an AUC of 0.63.

On the other hand, different strategies have been proposed to overcome class imbalance problems, which are divided into two categories: Algorithm Level and Data Level. Nowadays, Transfer Learning method has also been used for prediction problems with imbalanced nature of dataset. The motivation of introducing Transfer Learning with ensemble algorithms for telecommunication churn prediction data classification comes from two main reasons. Firstly, several ensemble algorithms were applied by researchers and did not show satisfactory classification performance, i.e. only showed marginal improvement in the churn prediction. Secondly, Convolution Neural Networks (CNNs) are known to achieve better performance on images or correlated grids like topology. Consequently, in this work, feature vectors are converted into images, and deep learning methods are applied to achieve better accuracy. Finally, it has been observed through various research studies that Transfer Learning [16-19] using fine-tuning of an already trained network may give improved performance, as compared with designing and training a CNN from scratch.

The proposed Transfer Learning and Ensemble-based Meta-Classification "TL-DeepE" exploits three ideas to develop an effective prediction system:

1. Firstly, it uses Deep CNNs as base classifiers for the ensemble classification. Moreover, both original and converted features are considered. In the case of the converted features, the original numerical features are converted into 2D image data first and then, are provided to the base classifiers.
2. Secondly, the proposed methodology exploits the quick learning and knowledge transfer abilities of "Transfer Learning" in CNNs. To perform Transfer Learning in the CNNs models, TF Slim library [20] is used.
3. Thirdly, we exploit the learning and discrimination abilities of the high-level Genetic Programming and AdaBoost based ensemble classifier. In this regard, the predictions of the Deep CNNs are appended to the original feature vector and thus are used to build a final feature vector for the high-level ensemble classifier.

The remaining sections of this manuscript are organized as follows: Section 2 provides details of the proposed TL-DeepE technique. It explains both the Transfer Learning idea as well as the GP-AdaBoost ensemble-based classification. Section 3 provides details of the telecommunication dataset and experimental setup used in this work. Results and comparative analysis are discussed in section 4. Finally, section 5 concludes the paper.

2. Proposed TL-DeepE Prediction System

The proposed TL-DeepE technique for telecom churn prediction uses the concept of meta-classification in addition to Transfer Learning (Figure 1). Input features are converted from one-dimensional vector to two-dimensional image format and are used to train multiple CNNs, which in turns give churn predictions (collectively called 'prediction space'). These predictions are then appended with the original feature vector, and classification is performed on this 'extended' feature set using GP-AdaBoost ensemble. Experimental evaluations are performed by using 10-fold cross-validation on two standard telecommunication churn prediction datasets i.e. from Orange and Cell2cell. The exploitation of sample space by AdaBoosting and enhancement of the learning and generalization by GP [21] makes the proposed TL-DeepE an effective method for churn prediction and classification.

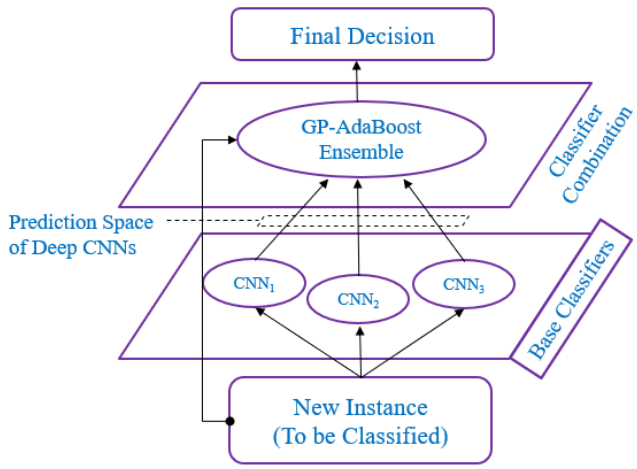


Figure 1: An Overview of the proposed TL-DeepE.

2.1 Data Pre-Processing

Telecom datasets have many concerns to address like missing values, non-numeric feature values, inconsistent feature scales, etc. It is, therefore, necessary to pre-process the data before applying a learning model. The datasets used, namely Cell2cell and Orange, have their individual characteristics. Initially, missing values from the dataset were handled by numeric and categorical feature vectors, respectively. Missing values from the dataset were inserted as column mean and mode for numeric and categorical feature vectors, respectively. This was done for the features which had not an overwhelming number of missing values. The features with more than 95% of missing values were removed from the dataset since it is assumed that they contribute little to the learning. Next, the categorical features are converted to numeric using the 'dummy variable' technique where different categories are split to form their own features. 1 or 0 is assigned as a value depending on whether the instance had the category as its original feature or not, respectively. Once the entire dataset is converted into a numeric format, we normalize it on a scale of 0 to 1 in order to scale the features uniformly.

2.2 Conversion of 1D Features to Images

A major aspect of this research work was to explore the possibility of converting the raw feature vector-based dataset to image form, with the belief that better classification accuracy can be achieved through this way using CNNs. The assumption behind this hypothesis is that CNNs are known to perform well with image data. Every single instance present in the dataset was converted to an individual representative image. To achieve this, the one-dimensional feature vector of each instance is converted into a two-dimensional matrix form, on the hit-and-trial basis, and it was found that the images obtained using the method of interpolation works well as far as the accuracy is concerned. Specifically, we construct a 2-dimensional matrix form of normalized pixels where each row is for each day and each column is for each type of behavior. Therefore, this method was used to convert all instances to image format and in turn, used for training and testing of CNN models. It is hypothesized that one can exploit the pixel-based differences between churner and non-churner instance images when it comes to CNNs learning.

2.3 Development of Individual Prediction Space using Transfer Learning

Training a CNN on a new dataset from scratch can be a very time consuming and resource intensive practice. The concept of Transfer Learning offers an alternative route to avoid learning from scratch. This concept introduces that instead of training from scratch, CNNs already trained on massive dataset like ImageNet should be fine-tuned (i.e. the weight updatation is performed) on the required dataset.

In order to obtain the decision space, three CNN models were used: AlexNet, Inception-ResNet-V2, and a custom 6-layer neural network model. These CNNs had been pre-trained on ImageNet dataset and Transfer Learning was used to fine-tune them on the Orange and Cell2cell datasets.

2.3.1 AlexNet

AlexNet is a famous CNN algorithm proposed by Alex Krizhevsky for ImageNet classification[22, 23] method. The architecture is composed of eight layers: first five are convolution layers with 4-pixel stride and next three are fully-connected and lead to the final layer, i.e. softmax. Softmax layer provides the probabilistic distributed decision over 1000 classes. Multinomial logistic regression objective is maximized by the network. Similarly, further layers have kernels smaller in size but an increased number of channels, such as 5x5x48 and 3x3x256. Fully-connected layers each have 4096 neurons and a final 1000-way softmax layer is contemporary at the end. The network uses data augmentation and dropout to reduce the effects of overfitting.

2.3.2 Inception-ResNet-V2

Inception family is a class of deep CNN architectures proposed by Google. Inception family architectures usually need more computational power as compared with AlexNet because of their more complex architecture. The concept behind Inception-ResNet-v2 [24] was to train deeper neural network architectures using skip connections [25, 26] and to automate the filter size selection. A single block in Inception ResNet architecture uses a filter bank with multiple sizes of filters (5x5, 3x3, 1x1), and apart from learning filter weights, the network also learns the filter sizes, which can work well for the given problem. On the ILSVRC image classification benchmark, this architecture demonstrated an accuracy of 95.3%, which falls among the top 5 of its category [27].

2.3.3 Custom-CNN Architecture

For comparison purposes, a custom deep neural network [28, 29] algorithm was also trained after hit-and-trial experimentation with different architecture parameters. The architecture was built in the layer patterns of conv-relu-pool-fc-softmax-class. It takes 32x32 input image, 3x3x6 filter with no padding at first Convolution layer and then 2x3 pooling, 5x5x10 at the second layer, 2x2 max-pooling, and final layer, softmax classification have 2 neurons.

These CNNs were originally pre-trained on ImageNet data and then were fine-tuned on the required Orange and Cell2cell datasets according to Transfer Learning task requirements. Once trained, the models were tested on the test set of images and their predictions were collected. These predictions were then appended to the original datasets to get extended feature space. The extended feature space was passed on for training and testing of the high-level GP-AdaBoost ensemble classifier.

2.3.4 Exploitation of Individual Prediction Spaces using GP-AdaBoost Ensemble Based Meta Classification

GP has greatly overcome different problems of clustering and regression. The capacity and flexibility of GP gives appropriate methods for classification and churn prediction. Similarly, AdaBoost, on the other hand, is a boosting technique that combines small classifiers to generate a high-level, stronger one using multiple programs at each class. GP-AdaBoost ensemble [30] incorporates the idea of boosting in evolving different GP programs. Every base classifier in a subsequent step is evolved multiple programs at each class to identify the 'hard samples' that were incorrectly classified by the preceding classifier. The methods at the earlier stage divide the data into two sets, i.e. training and testing. In Figure 2, a number of GP programs are evolved at given fixed Elite size in each class, whereas weighted sum is modified in an AdaBoost. Final prediction is made with the subsequent program by means of the high weighted sum output in each class. The optimized features achieved through GP are boosted by an AdaBoost method that eventually improved the prediction.

The Genetic Programming-AdaBoost [31] ensemble is used as a high-level meta-classifier on top of the base results achieved by the CNNs and Transfer Learning. The method extracts discriminative features using deep autoencoder and concatenates to original feature to form images. The extended feature vector uses the original feature vector concatenated with the base CNNs predicting space as input and outputs the final predictions.

3. Performance Evaluation

Generally, results of churn prediction systems are compared by standard measures like prediction accuracy. But in telecommunication datasets, accuracy is not sufficient to achieve the true picture of results due to an imbalanced distribution of the data. Therefore, the Area Under Curve (AUC) of Receiver Operating Characteristics (ROC) along with prediction accuracy is used as a performance measure to estimate the true functionality of the predictor. ROC [32] curve are generated on the basis of true positive rate (Sensitivity) and false negative rate (1-Specificity). Let's TP represent True Positives, FP False Positives, TN True Negatives, and FN False Negatives. The positive prediction (TP+FN) shows the true churners rate and negatives prediction (TN+FP) shows non-churners rate.

3.1 Prediction Accuracy

Prediction accuracy (Equation 1) gives a measure of the correct predictions against the total number of cases evaluated. The metric of prediction accuracy has been considered for the Transfer Learning task since the number of correct predictions from the neural network models would affect the overall ensemble classifier's result. The prediction accuracy is given below:

$$Prediction\ Accuracy = \frac{\text{TP+TN}}{TP+TN+FP+FN} \tag{1}$$

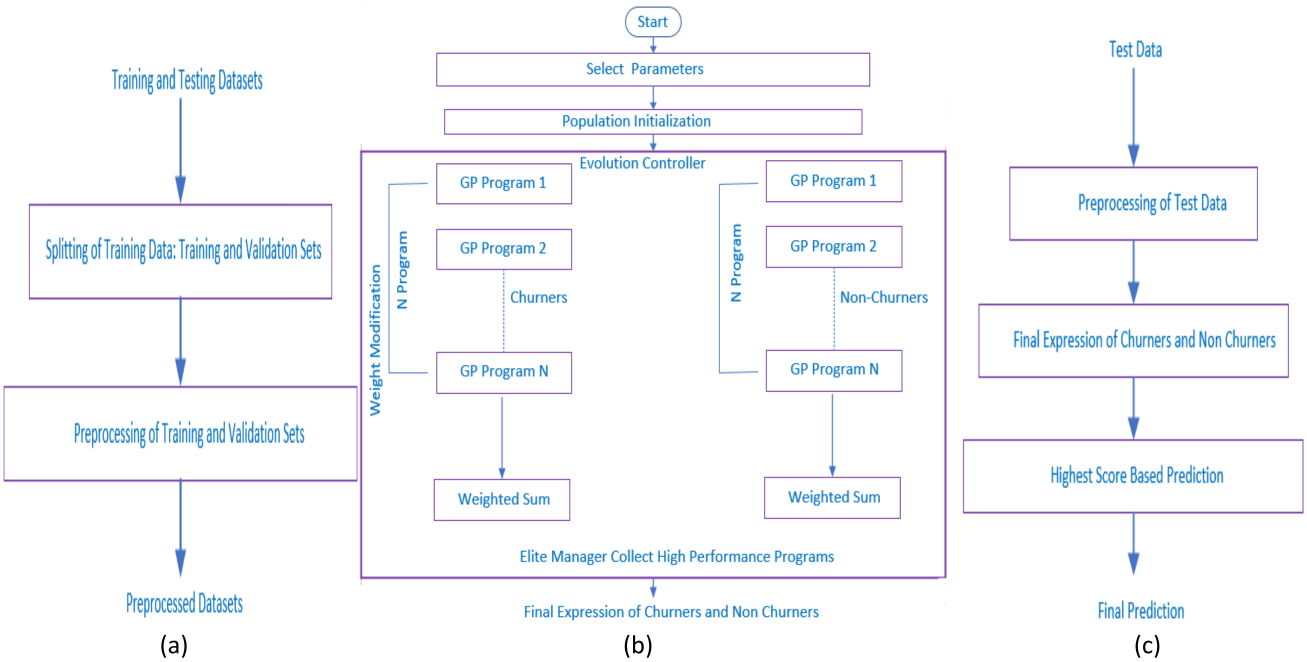


Figure 2: Block diagram of GP-AdaBoost approach: (a) Preprocessed datasets, (b) Final expression of Churner and non-Churners using Genetic Programming, and (c) Final churners prediction.

3.2 ROC Curve

The ROC curve and other related evaluation measurements are pertinent in two ways. Firstly, it provides a comparatively unbiased diagnostic ability of a binary classifier as compared to prediction accuracy, which can easily become biased in the case of imbalanced datasets. Secondly, it is helpful while comparing the performance of different methods, reported in the literature. Equations (2), and (3) show Sensitivity and Specificity parameters, while AUC is calculated as follow [33].

$$Sensitivity = \frac{TP}{TP+FN} \tag{2}$$

$$Specificity = \frac{TN}{TN+FP} \tag{3}$$

$$AUC = \int_0^1 \frac{TP}{TP+FN}\, d\, \frac{FP}{TN+FP} \tag{4}$$

4. Results and Discussion

Churn prediction system is assessed on the basis of its ability to correctly identify the churner in the telecom data. At present, most of the reported churn prediction systems are not capable of achieving accurate measurements or high accuracy due to serious complication in telecommunication data. The correct measurement of churned identification confidently protects telecommunication industries from critical losses. The proposed method for churn prediction shows convincing results against the telecommunication datasets and is better than other recent techniques.

4.1 Datasets

Two standard telecom datasets namely Orange and Cell2cell are used for the experimentation and analysis during this research work. The former one, i.e. Orange dataset from Telecom, UK, is made available to the researchers as part of KDD-Cup 2009 competition. The Cell2cell dataset, on the other hand, is from Duke University's Center for Customer Relationship Management. The Orange dataset has eighteen features with missing values and five features have just a single value, while Cell2cell dataset has no missing values. The symbolic feature values existing in telecommunication dataset are converted into the numerical forms. Attributes of these telecommunication datasets are described in Table 1.

Table 1: Characteristics of Telecom Datasets

Dataset	**Cell2cell**	**Orange**
Source	Duke University	KDD Cup 2009
Features	77	230
Samples	40000	50000
Features title	Well- defined	Undefined
Categorical variables	1	34
visioBehavior	Fair	Unfair
Positive samples	20000	46328
Negative samples	20000	3672
Missing features value	No	Yes

4.2 Dataset Splits and Overall Working of the Prediction System

As regards, the splitting of the data, there are various possibilities that the techniques can be applied throughout the classification phase. Furthermore, the distribution of dataset must be taken in the account before the whole process starts because some portion of the dataset will be used for training and the rest in testing the GP-AdaBoost ensemble. Therefore, both the datasets are divided into two subsets A and B after applying proper pre-processing techniques. Neural network models are trained on the subset A, which encompasses 60% of the entire dataset. These CNNs are used as the trained base learner in the GP-AdaBoost ensemble. Once trained, the model predicts the outcome of their learning on the subset B, i.e. encompassing 40% of the total data. Their test predictions on the subset B are completed, then these predictions are appended to subset B to form the extended feature set and saved in subset B'. The performance of the GP-AdaBoost on the subset B' is validated by using 10-fold cross-validation. The CNN models and the GP-AdaBoost classifier work independently, as described in Figure 3. The final prediction is provided by the GP-AdaBoost classifier.

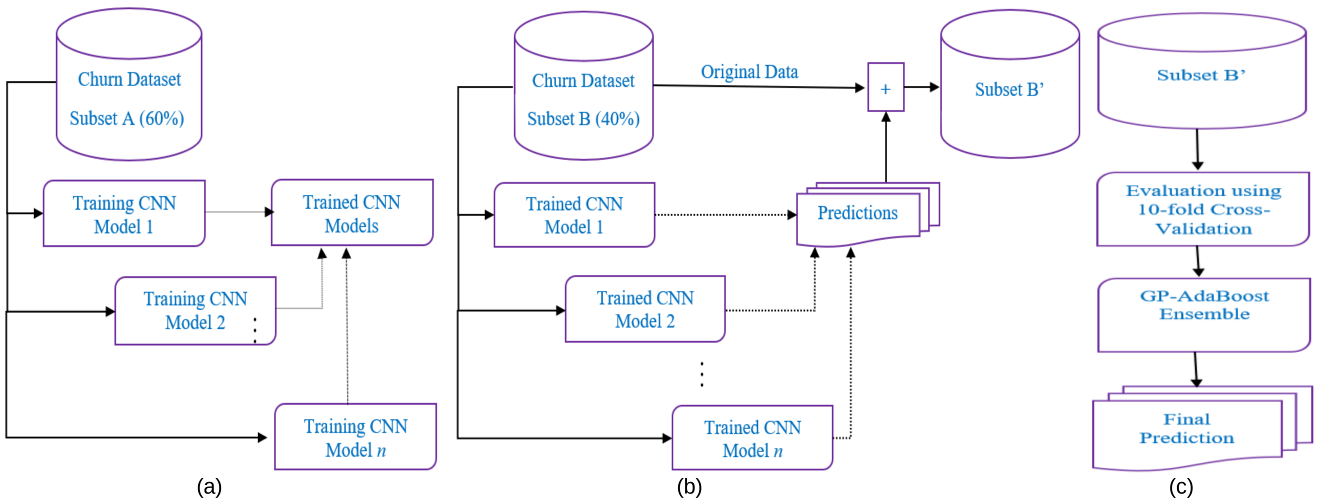


Figure 3: (a) Individual CNN Based Training, (b) Individual CNN Based Testing, (c) 10-fold GP-AdaBoost Based Training and Testing.

4.3 Churn Prediction Using Individual CNN Classifiers and Transfer Learning

The performance of the proposed prediction models (already pre-trained CNNs models that are fine-tuned) for Cell2cell and Orange datasets are summarized in Table 2. Inception-ResNet-V2 neural networks are pre-trained and then fine-tuned on telecom datasets. The Orange and Cell2cell datasets according to Transfer Learning provide prediction accuracies of 63.57 % and 68.01 %, respectively. It can be observed that the performance of the base learners is not satisfactory.

4.4 Ensemble Performance without Transfer Learning

The proposed method hypothesizes that Transfer Learning can improve the overall churn prediction and classification performance. In order to validate this idea, it is essential to evaluate the performance of TL-DeepE without Transfer Learning. In this regard, Table 3 shows the GP-AdaBoost ensemble performance without Transfer Learning on Orange and Cell2cell datasets. It is observed that the performance of the GP-AdaBoost ensemble without Transfer learning is not satisfactory and even lower than some of the individual base learners.

4.5 Ensemble Performance using Transfer Learning

Table 4 summarizes the results obtained by the proposed TL-DeepE (using Transfer Learning). In this case, the training and testing of the high-level meta-classifier are performed on two standard telecommunication datasets by evaluating on 10-fold cross-validation. The table depicts an average performance for 10 independent runs of the algorithm. Furthermore, AUC is shown in Figure 4(a). In Figure 4(b), comparisons are provided by evaluating the proposed method GP-AdaBoost ensemble both with and without Transfer Learning, on Orange and Cell2cell datasets. It is observed that the performance of the GP-AdaBoost ensemble in combination with Transfer learning is boosted and even higher than all of the deep individual base learners.

Table 2: Transfer Learning Based Results of Neural Network Models on Cell2cell and Orange Datasets.

Runs	Prediction Accuracy (%)					
	Cell2cell			Orange		
	AlexNet	Incep-RN-V2	Custom-NN	AlexNet	Incep-RN-V2	Custom-NN
Run 1	59.97	63.83	62.92	65.48	68.21	63.16
Run 2	59.45	63.65	61.08	66.19	67.74	63.04
Run 3	60.77	63.44	62.57	64.40	68.13	63.27
Run 4	60.53	63.25	62.60	64.89	67.78	62.95
Run 5	59.80	63.48	62.41	66.05	67.61	63.18
Run 6	60.96	62.98	61.35	65.86	67.94	63.07
Run 7	61.36	63.17	61.41	66.12	68.06	63.20
Run 8	60.20	64.27	61.20	65.29	68.49	62.84
Run 9	61.37	63.55	62.40	65.75	67.97	62.98
Run 10	61.74	64.05	62.37	66.10	68.24	62.80
Average	**60.62**	**63.57**	**62.03**	**65.61**	**68.01**	**63.05**

Table 3: GP-AdaBoost Ensemble Performance on Cell2Cell and Orange Datasets without Transfer Learning.

Runs	Cell2Cell		Orange	
	Prediction Accuracy (%)	AUC	Prediction Accuracy (%)	AUC
Run 1	60.30	0.64	62.16	0.69
Run 2	60.62	0.65	61.47	0.65
Run 3	61.22	0.64	62.81	0.67
Run 4	61.00	0.64	62.04	0.66
Run 5	60.33	0.62	63.19	0.67
Run 6	61.04	0.63	63.13	0.67
Run 7	60.90	0.63	62.48	0.66
Run 8	60.57	0.63	62.94	0.66
Run 9	61.18	0.64	62.03	0.66
Run 10	60.43	0.63	62.80	0.64
Average	**60.76**	**0.64**	**62.50**	**0.66**

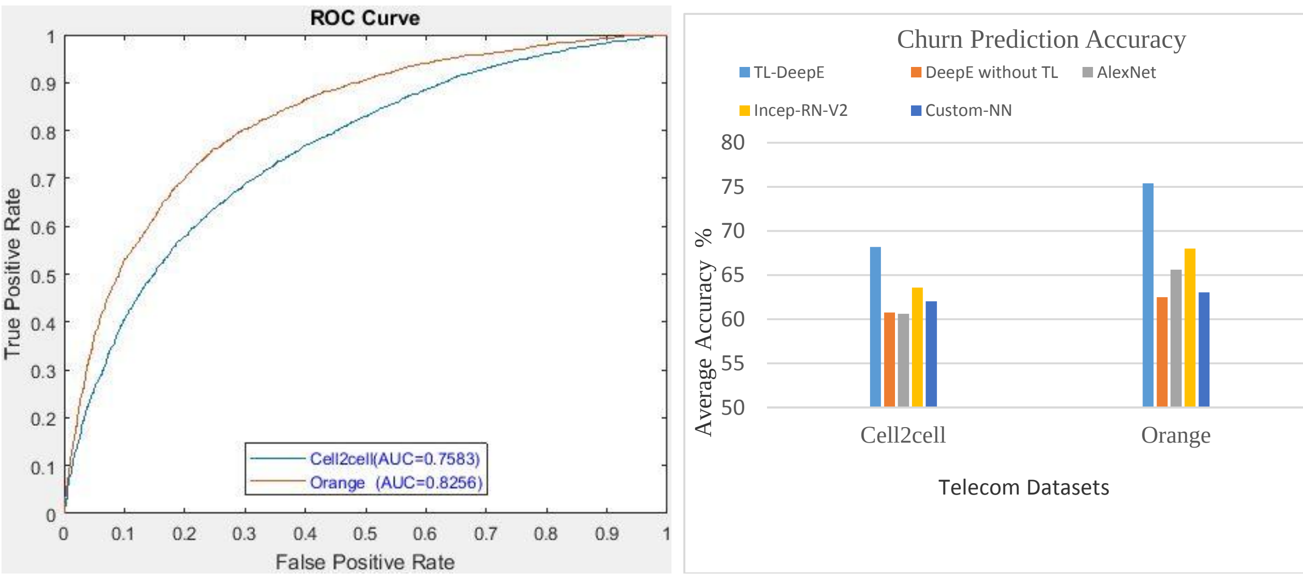
Table 4: Prediction Performance of the proposed TL-DeepE on Cell2cell and Orange Datasets.

Runs	Cell2Cell		Orange	
	Prediction Accuracy (%)	AUC	Prediction Accuracy (%)	AUC
Run 1	68.07	0.74	76.13	0.83
Run 2	67.86	0.74	75.90	0.83
Run 3	69.30	0.75	75.33	0.83
Run 4	68.65	0.74	75.51	0.83
Run 5	68.42	0.75	74.94	0.82
Run 6	67.38	0.73	75.11	0.83
Run 7	67.74	0.73	75.77	0.83
Run 8	68.40	0.74	75.50	0.83
Run 9	67.91	0.73	74.11	0.83
Run 10	67.84	0.73	75.51	0.83
Average	**68.16**	**0.74**	**75.38**	**0.83**

4.6 Comparative Analysis

Performance of the proposed churn prediction algorithm, i.e. TL-DeepE, is compared with previous techniques, on Cell2cell and Orange telecom datasets. Initially, the performance of the churn predictor like AlexNet, ResNet-V2, Custom-CNN, and then GP-AdaBoost ensemble classifier without Transfer Learning is evaluated on the telecommunication original churn prediction dataset. The poor performance of Alex Net, ResNet-V2, Custom-CNN neural networks is because of the already performed pre-tuning and little fine-tuning on telecom datasets. Hence a better prediction system is mandatory in order to avoid customer churn prediction complications.

The proposed system "TL-DeepE" in Figure 4 (b) shows that the integration of information from Transfer Learning with GP-Adaboost ensemble improves churn prediction performance over the complex datasets and surpasses the performance of all the earlier reported systems. The proposed system achieved the highest churn prediction accuracy of 75.4% and 0.83 AUC on Orange and 68.2%, 0.74 AUC on Cell2Cell telecommunication datasets, respectively. This shows its strength to perform on complicated nature of the telecom data.



(a) (b)

Figure 4. (a) ROC curve of the test results of GP- AdaBoost on Cell2cell and Orange dataset and (b) Comparison of Churn predictors.

5. Conclusion

A novel TL-DeepE system is proposed to predict potential churners, which is very important for the competitive telecom industries. The size and dimensionality of the telecom data are extremely high and involves a large computational power for churn prediction. The intelligent system makes use of multiple CNN architectures and GP-AdaBoost ensemble with the exploitation of Transfer Learning concepts. The individual prediction of the base classifiers is then appended the original feature vectors of the dataset to form extended feature vectors for each of the datasets. Finally, using a 10-fold cross-validation technique, a GP-AdaBoost ensemble as meta-classifier is trained and evaluated to obtain the correct predictions.

The proposed TL-DeepE churn prediction system has shown its effectiveness in predicting churners in standard telecommunication datasets. TL-DeepE has demonstrated churn prediction accuracy of 75.4% and 0.83 AUC on Orange dataset. On the other hand, its accuracy was 68.2% and AUC 0.74 on Cell2Cell dataset. It is observed that due to the challenging nature of the churn prediction task, choice of suitable features and the exploitation of Deep Learning, Transfer Learning, and GP-AdaBoost meta-classification appear more effective for modeling churn prediction in the telecommunication industry. The proposed method can be potentially effective in addressing the concerns and complications of the telecommunication industry.

Acknowledgment:

This research was supported by HEC NRPU Research Grant (No. 20-3408/R&D/HEC/14/233) Pakistan, and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1A2B2005065).

References:

1. Kim, M.-K., M.-C. Park, and D.-H. Jeong, *The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services.* Telecommunications policy, 2004. **28**(2): p. 145-159.
2. Verbeke, W., et al., *New insights into churn prediction in the telecommunication sector: A profit-driven data mining approach.* European Journal of Operational Research, 2012. **218**(1): p. 211-229.

3. Yabas, U. and H.C. Cankaya. *Churn prediction in subscriber management for mobile and wireless communications services*. in *Globecom Workshops (GC Wkshps), 2013 IEEE*. 2013. IEEE.

4. Ahmad, N., et al. *Computer vision based room interior design*. in *Eighth International Conference on Machine Vision (ICMV 2015)*. 2015. International Society for Optics and Photonics.

5. Reichheld, F. and C. Detrick, *Loyalty: A prescription for cutting costs.* Marketing Management, 2003. **12**(5): p. 24-24.

6. Khaleghi, B. *What makes predicting customer churn a challenge*. 2017; Available from: https://medium.com/@b.khaleghi/what-makes-predicting-customer-churn-a-challenge-be195f35366e.

7. Lessmann, S. and S. Voß, *A reference model for customer-centric data mining with support vector machines.* European Journal of Operational Research, 2009. **199**(2): p. 520-530.

8. Suryadi, K. and S. Gumilang. *Actionable decision model in customer churn monitoring based on support vector machines technique*. in *9th Asia Pacific Industrial Engineering and Management Systems Conference, Bandung, Indonesia*. 2008.

9. Idris, A. and A. Khan, *Churn prediction system for telecom using filter–wrapper and ensemble classification.* The Computer Journal, 2016. **60**(3): p. 410-430.

10. Raschka, S. *Stacking Classifier*. 2017; Available from https://rasbt.github.io/mlxtend/user_guide/classifier/StackingClassifier.

11. De Bock, K.W. and D. Van den Poel, *An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction.* Expert Systems with Applications, 2011. **38**(10): p. 12293-12301.

12. Neslin, S., *Cell2Cell: The churn game.* Cell2Cell Case Notes. Hanover, NH: Tuck School of Business, Dartmoth College, 2002.

13. Idris, A., A. Khan, and Y.S. Lee, *Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification.* Applied intelligence, 2013. **39**(3): p. 659-672.

14. Xie, Y., et al., *Customer churn prediction using improved balanced random forests.* Expert Systems with Applications, 2009. **36**(3): p. 5445-5449.

15. Idris, A., A. Khan, and Y.S. Lee. *Genetic programming and adaboosting based churn prediction for telecom*. in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*. 2012. IEEE.

16. Pan, S.J. and Q. Yang, *A survey on transfer learning.* IEEE Transactions on knowledge and data engineering, 2010. **22**(10): p. 1345-1359.

17. Yosinski, J., et al. *How transferable are featured in deep neural networks?* in *Advances in neural information processing systems*. 2014.

18. Qureshi, A.S. and A. Khan, *Adaptive Transfer Learning in Deep Neural Networks: Wind Power Prediction using Knowledge Transfer from Region to Region and Between Different Task Domains.* arXiv preprint arXiv:1810.12611, 2018.

19. Qureshi, A.S., et al., *Wind power prediction using deep neural network based meta-regression and transfer learning.* Applied Soft Computing, 2017. **58**: p. 742-755.

20. Guadarrama, N.S.a.S. *Tensor Flow-Slim Image Classification Model Library*. 2016; Available from https://github.com/tensorflow/models/tree/master/research/slim.

21. Koza, J.R. *How Genetic Programming Works*. 2007; Available from http://www.genetic-programming.org.

22. Russakovsky, O., et al., *Imagenet large scale visual recognition challenge.* International Journal of Computer Vision, 2015. **115**(3): p. 211-252.

23. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.

24. Sin, K. *Transfer Learning in Tensorflow using a Pre-trained Inception-Resnet-v2 Model*. 2017; Available from https://kwotsin.github.io/tech/2017/02/11/transfer-learning.html.

25. Szegedy, C., et al. *Inception-v4, inception-resnet and the impact of residual connections on learning*. in *AAAI*. 2017.

26. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

27. Tsang, S. *Winner in ILSVRC 2016 (Image Classification)*. 2018; Available from https://towardsdatascience.com/review-trimps-soushen-winner-in-ilsvrc-2016-image-classification-dfbc423111dd.

28. The team, T.D. *Convolutional Neural Networks (LeNet)*. 2015; Available from: http://deeplearning.net/tutorial/lenet.html. .

29. Castrounis, A. *Artificial Intelligence, Deep Learning, and Neural Networks, Explained*. 2016; Available from http://www.kdnuggets.com/2016/10/artificial-intelligence-deep-learning-neural-networks-explained.html.

30. Freund, Y. and R.E. Schapire, *A Short Introduction to Boosting": Journal of Japanese Society for Artificial Intelligence, 14 (5): 771-780.* 1999.

31. Khan, A., A. Sohail, and A. Ali, *A New Channel Boosted Convolutional Neural Network using Transfer Learning.* arXiv preprint arXiv:1804.08528, 2018.

32. Zou, K.H., A.J. O'Malley, and L. Mauri, *Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models.* Circulation, 2007. **115**(5): p. 654-657.

33. Burez, J. and D. Van den Poel, *Handling class imbalance in customer churn prediction.* Expert Systems with Applications, 2009. **36**(3): p. 4626-4636.