Accuracy Performance Degradation in Image Classification Models due to Concept Drift

Manzoor Ahmed Hashmani[1], Syed Muslim Jameel[2], Hitham Alhussain[3], Mobashar Rehman[4], Arif Budiman[5]

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Sri Iskandar, Malaysia[1, 2, 3]

Universiti Tunku Abdul Rahman, Kampar, Malaysia[4]

University of Indonesia, Jakarta, Indonesia[5]

Abstract—**Big data is playing a significant role in the current computing revolution. Industries and organizations are utilizing their insights for Business Intelligence by using Deep Learning Networks (DLN). However, dynamic characteristics of BD introduce many critical issues for DLN; Concept Drift (CD) is one of them. CD issue appears frequently in Online Supervised Learning environments in which data trends change over time. The problem may even worsen in a BD environment due to the veracity and variability factors. The CD issue may render the DLN inapplicable by degrading the accuracy of classification results in DLN which is a very serious issue that needs to be addressed. Therefore, these DLN need to quickly adapt to changes for maintaining the accuracy level of the results. To overcome classification accuracy, we need some dynamical changes in the existing DLN. Therefore, in this paper, we examine some of the existing Shallow Learning and Deep Learning models and their behavior before and after the Concept Drift (in experiment 1) and validate the pre-trained Deep Learning network (ResNet-50). In future work, this experiment will examine the most recent pre-trained DLN (Alex Net, VGG16, VGG19) and identify their suitability to overcome Concept Drift using fine-tuning and transfer learning approaches.**

Keywords—Pre-trained networks; deep learning; concept drift; fine tuning; transfer learning

I. INTRODUCTION

Machine Learning (ML) is being actively investigated by researchers for the last few decades. However, the need for utilizing these ML models in production has introduced new issues and problems for the research community. Interestingly, the term Industrial Revolution 4.0 has not only changed the traditional business process, but also the traditional Machine Learning approach [1]. The essence of these new approaches is to harness the power of Big Data. Among other types of Big Data (text, audio, sensor data, images, etc.), images are significant. Images are being utilized for prediction and classification in health, education and industries (automobiles, agriculture, etc.) employing the Image Classification Models (SVM, ELM, CNN, OSELM, ACNNELM, etc.) [2]. Industrial Revolution 4.0 demands more robust and scalable Image Classification models. IR4.0 related applications need to classify input data stream in real time. Therefore, instead of batch processing (offline training), these models must learn online learning scenarios (learning with classification). However, in Online Learning scenarios, the statistical properties of images may vary at different time steps which

substantially decrease the performance in terms of accuracy in Image Classification Models. This phenomenon is also known as Concept Drift issue.

II. RELATED WORK

In literature, several studies discussed and proposed solutions for the Concept Drift issue in online machine learning scenario that handled CD in the non-imaging dataset. However, handling of Concept Drift in Image Classification models is rarely reported. Shallow ML classification models (e.g., Extreme Learning Machine (ELM), Support Vector Machine (SMV), Multi-Layer Perception Neural Network (MLP NN), Hidden Markov Model, etc.) handle classification and regression problems efficiently in structured data [3] and are not feasible to handle the large Image datasets [4]. However, Deep Learning algorithms are a better suited to handle complex data streams and extract value with higher accuracy as compared to the conventional approaches. However, the issue of CD can be handled in online learning through new data adaption models/ modes/ concepts. Therefore, truly autonomous, self-maintaining and adaptive Image Classification models are needed. Some studies focused on CD and adaptation of systems during online sequential environment and urged researchers to further investigate in this direction [5].

A recent study [1] presented a comprehensive view of all existing Image Classification models to handle the Concept Drift issue. This study summarizes the available literature in the following categories:

- Non-Adaptive Shallow Learning Approaches
- Adaptive Shallow Learning Approaches
- Non-Adaptive Deep Learning Approaches
- Adaptive Deep Learning Approaches
- Adaptive Hybrid Deep Learning Approaches

Moreover, this study also proposes a Fully Adaptive Image Classification approach using Meta-cognitive principles by using an ensemble classifier approach. However, the selection of an optimized classifier is very critical. Therefore, in this paper, we have conducted several experiments on several Machine Learning models (including pre-trained Deep Learning models) to validate their better performance in the CD scenario.

III. RESEARCH METHODOLOGY

This research study examines the multiple Concept Drift scenario (Real, Virtual and Hybrid) and its behavior with several Shallow Learning (SVM, ELM, OSELM) and Deep Learning models (CNN, ResNet-50) in online image stream scenarios. MNIST and CIFAR 10 are considered as benchmark datasets [6] for grey scale (2 channels) and RGB (3 channels) images [7]. In this study we carried out two different experiments to understand the behavior of Machine Learning models due to Concept Drift. They are defined below:

Experiment 01: To identify the effect of Concept Drift on Shallow and Deep Learning Models.

Experiment 02: To mitigate the adverse effect of Concept Drift through the pre-trained network.

Experiment 01 is dependent on two major parts. Firstly, the Machine Learning (ML) models (SVM, ELM, OSELM, and CNN) were trained and tested using MNIST dataset (to illustrate the normal behavior of those Machine Learning models for greyscale images). After that, we simulated the Real, Virtual and Hybrid CD and analyzed their effects on SVM, ELM, OSELM, and CNN. Secondly, we did the same experiment for CIFAR 10 dataset to analyze the CD impact on ML due to RGB images.

In Experiment 02, we used the pre-trained network ResNet-50. Firstly, the ResNet-50 was evaluated by using several natural images. The testing accuracy of these tests was recorded. However, we used CNN for feature extraction and SVM for the classification task. Three categories were used (Airplane, Ferry and Laptop among 101 categories) for simplicity (airplane =800 images, ferry=67 images, laptop=80). These classes were balanced to 67 each because they contained the class imbalance problem (check what if we don't balance these classes). We just retrained the last layer of ResNet-50 with additional three categories (airplane, ferry, and laptop). Note that these layers were trained using 1000 classes from ImageNet dataset. We divided the dataset into 30% training and 70% validation (through randomize technique). Caltex 10 image pixel value is 300x300x3 whereas ResNet-50 takes an input value of 224x224x3. Therefore, we used augmented image datastore (it adjusts image size as per input size and gray-scale into RGB). We can get the deep layer features by using activation (it is better to extract features through the before layer of classifier layer). A minibatch of size 32 was used for getting an optimized GPU utilization. It is better to arrange activation output as a column to obtain more/higher GPU optimization. We used Stochastic Gradient Descent (SGD) for cost optimization (a vector was extracted from CNN for this feature), the SVM classifier (fitccoc in Matlab) was trained by using CNN features. The mean accuracy was evaluated and displayed by using the Confusion Matrix. Each layer of CNN makes some contribution to the input image (by applying/updating weights or adjusting the activation function). We can see by visualizing the network filter`s weight.

A. Dataset

In Machine Learning, MNIST is recognized as the benchmark dataset for greyscale images [6][7][8]. The MNIST dataset contained 70,000 handwritten images (28x28 pixel) with 10 target classes [9]. However, for color images, CIFAR 10 [10] and CALTECH 101 [11] datasets are considered as benchmark for classification problem. Caltech 101, one of the most widely cited and used image datasets, was collected by Fei-Fei Li, Marco Andreetto, and Marc 'Aurelio Ranzato.

B. Models

In Experiment 01, this study analyzed the performance accuracy for Support Vector Machine (SVM) [12], Extreme Learning Machine (ELM) [13], Online Sequential Extreme Learning Machine (OSELM) [14] and Convolutional Neural Network (CNN) [15]. However, Experiment 02 used network ResNet-50 [16] to investigate the behavior of pre-trained networks before and after CD scenario. In the ResNet-50 networks, initial layers are based on CNN structure, which is used to extract its own feature unlike SVM (which extracted hand-crafted features i.e. SURF, HOG, and SPARSE). The initial layers of CNN extract the basic image feature i.e. blob, edges, etc. ResNet-50 model is already trained from 1 million images of 1,000 classes using ImageNet [17] dataset (e.g., AlexNet, GoogleNet, VGG-16, and VGG-19). It uses CNN part for image feature extraction (It requires input image size 224x224x3). However, we used Support Vector Machine (SVM) as classification layer. We trained SVM classifier on three classes (discussed in methodology) from CALTECH 101 dataset.

C. Software and Hardware

Both experiments were conducted on a single node with parallel processing graphic card (G-FORCE NVIDIA GPU TITAN XP) containing 3748 cores with 32 GB RAM. MATLAB Statistical Machine Learning Toolbox, Deep Learning Toolbox, and ResNet-50 API were utilized to perform these experiments.

IV. RESULTS AND DISCUSSION

In this study, we conducted two different experiments to validate the performance degradation (in terms of accuracy) due to Concept Drift in SVM, ELM, OSELM, and CNN using Image Dataset.

Initially, ELM, SVM, CNN, and OSELM were trained by using MNIST (784 input pixels and 10 classes) dataset. Later, these trained models validated performance accuracy by testing dataset (a subset of MNIST). The results in Table 1 show that the accuracy in a normal condition is acceptable. The CNN, however, performs well among all other existing models. To simulate the Virtual Concept Drift, an additional feature Histogram of Gradient (HoG) was added to the testing dataset. After testing, the trained models (SVM, ELM, CNN and OSELM) degraded their classification accuracy by almost more than 20% (as shown in Tables 1 and 2).

To simulate the Real Concept drift scenarios, three different data streams (with a change in class boundary) were created. The existing trained models (SVM, ELM, CNN and OSELM) were tested using these three Data Streams (Data1, Data2 and Data3) (shown in Table 3). After testing from Data1, the performance accuracy did not change significantly (because Data1 is the same data stream on which these models were

trained). However, the classification accuracy decreased substantially after Data2 and Data3 (shown in Table 4).

Hybrid Concept Drift is the condition in which both Virtual and Real drifts take place simultaneously. Therefore, to simulate the Hybrid Drift, the properties both Virtual and Real drifts were added to the testing dataset (as shown in Table 5). The already trained models were tested before and after the Hybrid Drift conditions. The results indicated that, after Hybrid Drift, the decrease in the accuracy performance was more than the accuracy degradation in Virtual Drift and Real Drift (shown in Table 6).

Next experiment was conducted to analyze the behavior of SVM, ELM, CNN and OSELM models for 3 channels (RGB, color) images. In this experiment, only hybrid drift was simulated in this experiment because it contained the properties of both types of Concept Drift. Initially, the models were trained using Data1 (which is only RG channels). In order to maintain the input image size, we added 1 channel of Red and 2 channels of Green. The testing results were very nominal after training from less amount of dataset (CIFAR 10) (shown in Table 8). Thereafter, additional 1 channel (blue) was added to simulate the Hybrid Drift in testing data (shown in Table 7). After the occurrence of drift scenario, the accuracy performance of all models became worse to such an extent that even the accuracy of the CNN model was also degraded by almost 50%.

TABLE I.   MNIST AND MNIST-HOG DATASETS TO SIMULATE VIRTUAL CD SCENARIO

Dataset	Input	Output
MNIST	784	10 classes (0-9)
MNIST-HoG	784x81=865	10 classes (0-9)

TABLE II.   TESTING ACCURACY AFTER VIRTUAL DRIFT

Models	Accuracy before CD (MNIST)	Accuracy after CD (MNIST-HoG)
ELM	92.7%	70.2%
SVM	89.0%	68.8%
CNN	99.7%	78.4%
OSELM	95.2%	73.2%

TABLE III.   TMNIST DATA STREAMS TO SIMULATE REAL CD SCENARIO

Dataset	Input	Output
MNIST (Data1)	784	10 (0-9)
MNIST (Data2)	784	20 (A-J, 0-9)
MNIST (Data3)	784	4 (6-9)

TABLE IV.   TESTING ACCURACY BEFORE AND AFTER REAL DRIFT

Models	Before CD (Data1)	After CD (Data2)	After CD (Data)
ELM	92.12%	65.76%	69.56
SVM	88.67%	70.68%	68.78
CNN	99.76%	76.76%	73.45
OSELM	94.56%	72.34%	69.67

TABLE V.   MNIST DATA STREAMS TO SIMULATE HYBRID CD SCENARIO

Dataset	Input	Output
MNIST (Data1)	784	10 (0-9)
MNIST (Data2)	784+81=865	4 (6-9)

TABLE VI.   TESTING ACCURACY BEFORE AND AFTER HYBRID DRIFT

Models	Before CD (Data1)	After CD (Data2)
ELM	92	60.32
SVM	89	59.46
CNN	100	72.87
OSELM	95	69.59

TABLE VII.   CIFAR 10 DATA STREAMS TO SIMULATE HYBRID CD SCENARIO

Dataset	Input	Output
CIFAR-10 (Data1)	32x32x3 (3072) {RG+G}	10
CIFAR-10 (Data2)	32x32x (2+1) = (3072) {RG + B}	5

TABLE VIII.   TESTING ACCURACY BEFORE AND AFTER HYBRID DRIFT

Models	Before CD (Data1)	After CD (Data2)
ELM	70.49	25.26
SVM	77.68	34.75
CNN	97.8	49.50
OSELM	91.28	39.67

In experiment 2, a pre-trained network (ResNet-50) was validated in the Real Concept Drift condition. However, it can be noticed that the accuracy performance was more than all other existing models (shown in Table 9).

TABLE IX.   MEAN ACCURACY OF RESNET-50 WITH SVM CLASSIFIER

Models	Dataset	Mean Accuracy Before CD	Mean Accuracy After CD
ResNet-50	Caltech 101	98.58	76.81

V.   CONCLUSION

It can be safely concluded from the results of this experimental study, that the Image Classification model degrades its performance (in terms of accuracy) due to Concept Drift. However, the Hybrid drift causes more accuracy degradation then Real or Virtual Drift. Moreover, the complexity of image dataset is also directly proportional to the accuracy degradation after Concept Drift (e.g., MNIST image's accuracy degradation is less than CIFAR 10). Interestingly, the CNN model showed higher accuracy than others in most of the experiments because it has extracted and not hand-crafted features (because it is extracted not hand-crafted features). However, the accuracy of pre-trained network (ResNet-50) is better than that of CNN. Fundamentally, ResNet-50 also uses CNN for feature extraction, whereas its feature extraction layers are already trained using ImageNet dataset (containing 1 million images of 1000 classes). Therefore, through these experiments, we can conclude that the pre-trained network

offers a better solution for handling Concept Drift at the classifier level. Nonetheless, there is a need of developing a dynamic adaptation approach (which will work along with classifier) to adapt new features of Image Data stream (in online learning scenario). In future work, this study will be extended towards the training and validation of pre-trained Deep Learning models (e.g., ResNet-50, AlexNet, VGG16, VGG19) in certain Concept Drift scenarios. For that reason, multiple CD scenarios will be added explicitly (by adding the new features and classes in input image dataset) to investigate the performance accuracy of other Deep Learning networks in CD environment. This experiment will help in understanding and analyzing the performance of pre-trained Deep Learning networks in a variety of CD scenarios.

REFERENCES

[1] Jameel, Syed Muslim, et al. "A Fully Adaptive Image Classification Approach for Industrial Revolution 4.0." International Conference of Reliable Information and Communication Technology. Springer, Cham, 2018.

[2] Najafabadi, M M,. Villanustre, F,. Khoshgoftaar, T M,. Seliya, N,. Wald, R,. Muharemagic, E. "Deep Learning Applications and Challenges in Big Data Analytics", Journal of Big Data 2(1), 1 (2015).

[3] Kuncheva L.I. "Classifier Ensembles for Changing Environments". In: Roli F., Kittler J., Windeatt T. (eds) Multiple Classifier Systems. MCS 2004. LNCS, vol. 3077, pp. 1-15. Springer, Berlin, Heidelberg (2004).

[4] Kuncheva L.I. "Classifier Ensembles for Detecting Concept Change in Streaming Data". Overview and Perspectives. 2nd Workshop SUEMA, pp. 5-10 (2008).

[5] Zliobaite, I., Bifet, A,. Pechenizkiy, M,. Bouchachia, A.. "A Survey on Concept Drift Adaptation". ACM Computer Survey 46(4), pp. 1-37 (2014).

[6] Budiman, A., Fanany, M I., Basaruddin C.: Adaptive Online Sequential ELM for Concept Drift Tackling. Computational Intelligence and Neuroscience 2016 (20), 2016.

[7] Corrigan, Owen. An Investigation Into Machine Learning Solutions Involving Time Series Across Different Problem Domains. Diss. Dublin City University, 2018.

[8] Papernot, Nicolas. "Adversarial Examples in Machine Learning." (2017).

[9] Y. LeCun, C. Cortes. MNIST handwritten digit database [online] (2010).

[10] Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams. "Practical bayesian optimization of machine learning algorithms." Advances in neural information processing systems. 2012.

[11] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[12] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

[13] Huang, Guang-Bin; Zhu, Qin-Yu; Siew, Chee-Kheong (2006). "Extreme learning machine: theory and applications". Neurocomputing. 70 (1): 489–501.

[14] G.-B. Huang, N.-Y. Liang, H.-J. Rong, P. Saratchandran, N. Sundararajan, On-line sequential extreme learning machine; M.H. Hamza (Ed.), IASTED International Conference on Computational Intelligence, IASTED/ACTA Press (2005), pp. 232-237

[15] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013.

[16] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[17] Russakovsky, O., Deng, J., Su, H. et al. Int J Comput Vis (2015) 115: 211. https://doi.org/10.1007/s11263-015-0816-y.