SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

# SAFE-SEMI SUPERVISED FRAME WORK THROUGH MACHINE LEARNING

**S.Jagadeesh**
Professor, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India, jaga.ssjec@gmail.com

**Aerva Bhargavi**
U.G. Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India

**Ruchitha Reddy Kuntla**
U.G. Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India

**Farheen Begum**
U.G. Student, Department of Electronics and Communication Engineering, Sridevi Women's Engineering College, Hyderabad, India

***Abstract***

Semi-supervised learning (SSL) is a machine learning algorithm totally depends on manifold regularization technique. We have different approaches in manifold regularization but we mostly use graphical approach. Though, the execution of Semi-supervised learning (SSL) depends on the construction of manifold graph and the safety degrees of unlabeled samples. As the SSL machine learning totally depends on the manifold graph (MR) and degrees of unlabeled samples i.e, those are usually constructed before the classification and fixed during the process of classification learning, then the results are independent with the classification. Focusing on the above problems, we proposed a unified algorithm called adaptive safe semi-supervised learning frame work. Here in this method construction of manifold graph and calculates the safety degrees of unlabeled samples. We will optimize the weights of manifold graph and degree of unlabeled samples in learning process and then it calculates in advance. At last, we develop and implement a adaptive safe semi-supervised extreme machine learning. Here the performance of this algorithm is effective, reliable and we get more accuracy as compared to others.

**key words:** Semi-supervised learning (SSL), extreme learning machine, adaptive safety degree, adaptive graph, manifold regularization (MR).

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

## 1.Introduction

Semi-supervised learning (SSL) has been successful in both theory and application field. The main reason is that obtaining marker samples are very hard and expensive. In many practical solved problems, it is very easier and less costly to collect unmarked samples. We all know that SSL has been widely applied in various fields. Generally, 5SL uses various assumptions to establish the relation between unlabeled and labeled samples, such as manifold graph assumptions. For example, PK Mallapragada proposed Boosting for Semi-Supervised learning. Here we learned more about pattern recognition and about machine learning. The most previous studies have totally focused on the designing special algorithms. The accuracy of unlabeled examples have more as compared to before. A well performing graph is created, that clears a way for subsequent classification, it can help in improvement of classification process. A well performing graph is created, that clears a way for subsequent c1assification, it can help in improvement of classification process. Otherwise, it may lead to the damage the performance. As graph is created in advance and kept fixed during the learning process. The performance of graph is impossible to check virtually. Some parameters need to be adjusted in the manifold graph. As parameter selection is still an efficient solution that has not been solved, it is very difficult to build a good performance graph before c1assification.It sets up another hurdle for constructing for MR in advance. As we know existing improvements of MR either to select regularization parameters or to improve efficiency. At present some studies have shown that unlabeled samples may be at risk.

Unlabeled samples cannot be safely used, it will limit the applications. It is necessary to design a safe semi supervised learning method that never performs worst, than corresponding SL method using labeled sample inputs. Now -a-days many supervised methods have been proposed. Extreme machine learning (ELM) was proposed by Huang et al. Unfortunately, this ELM is not possible for unlabeled samples. To control this problem we introduced semi supervised machine learning(SSL).This is applicable for both labeled and unlabeled samples.in this paper, we propose a unified adaptive safety semi-supervised learning framework

## 2. Literature Review

### Regularization of Manifold graph:

A Geometric *Framework Learning algorithms from Labeled and Unlabeled data (M Belkiii, 2006*). In this we propose different learning algorithms depending on new regularizations which make use of the marginal distribution. Here the core point is semi supervised framework that includes labeled data and unlabeled data for a general purpose learner. Few graph learning algorithms and few standard methods can be used as special cases. In this we use properties of kernel Hilbert methods for new theorems that provide theoretical base for every algorithm. Here as a result, acquisition of natural output and are also able to grasp both the transductive and truly semi-supervised outputs. Here are the experimental affirmations showing that semi-supervised algorithms can use unlabeled data successfully. Eventually a short discussion of unsupervised and supervised learning in a framework.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

**The Related Theory**

**Semi Boost (Boosting framework):**

For Semi-Supervised Learning (PK Mallapragada, 2009) Semi-supervised learning has influenced a important amount of observation in pattern recognition and machine learning. Previous studies have concentrated on designing special algorithms to usefully utilize unlabeled data in coincidence with labeled data. Our aim is to upgrade the classification accuracy of any supervised learning algorithm by utilizing the unlabeled examples. We label this as semi-supervised improvement problem, to differentiate the proposed approach from the existing approaches. Hence, we design a Meta semi-supervised learning.

**Survey on semi-supervised learning algorithm (JE van Engelen, 2020):**

Semi-supervised learning is a branch of machine learning interested in using both labeled and unlabeled data together to execute specific tasks. Functionally locate in the middle of supervised and unsupervised learning, it provides control over the large amounts of unlabeled data accessible in most of the used cases in combination with smaller sets of labeled data. In the past few years, experimentation in this area has followed the common trends in machine learning, by considering closely at neural network based models. The compositions on this topic has also been enlarged in volume and scope, and a spectrum on it theoretically and applications. However, no survey exists to explain all this knowledge, this slow downs the ability of researchers. Here we present all the updates of semi-supervised learning methods and its recent advances. We concentrated on semi-supervised classifications, here large majority of semi-supervised learning researches occur.

**Semi-supervised learning handwritten digit identification using few labeled data (S Van Vaerenbergh):**

We come up with a novel semi-supervised extractor for handwritten digit identification issues that are established on the expectation that any digit can be acquired as a little modification of another adequately close digit. If we take a number of labeled and unlabeled images, it is manageable to control the class membership of every unlabeled image by generating a series of alike image transformations that connect it to labeled images. At last the experimental results on the Modified national Institute of Standards and Technology indicates that the proposed classifier surpass current state-of-the-art techniques.

**Combination of Manifold Graph Regularization (B Geng, 2012):**

We come up with an automated estimation of the intrinsic manifold for general semi-supervised learning (SSL) issues. Unfortunately, it is not insignificant to describe an increment function to acquire optimal hyper parameters. Generally, cross validation is applied, but it does unnecessarily scale up. Thus, we develop an ensemble manifold regularizations framework. Further, we set up the convergence property of EMR matrix, Comprehensive experiments of

DOI: 10.5281/zenodo.6877693

both artificial and real data sets manifest the usefulness of the proposed framework.

**3. Finding**

**Semi-supervised learning (SSL)**

Extreme Machine Learning (ELM)is a very good learning tool for pattern recognition and machine learning. It randomly generates the hidden layer bias and input weights. When compared to neural networks algorithms, ELMs have advantages of low computational cost, good versatility and simple structure, ELM overcomes the drawbacks of neural networks. This ELM tool gives less amount of accuracy. To overcome this problem we have introduced adaptive safe semi supervised algorithms to get more accuracy.

An approach to machine learning that is the combination of a small amount of labeled data and large amount of unlabeled data during training is called as semi-supervised learning. It falls between two types of learning, one is unsupervised learning and supervised learning. Supervised learning requires hand labeling, by ML specialists or data scientists learning. To overcome drawbacks of supervised and unsupervised algorithms, this semi-supervised learning concept is proposed.

**Implementation**

The main aim is to compare the accuracy of existing and proposed systems. Here the usage of new programming language python to develop the code in easy way is used, code is developed in tkinter message box. The implementation we have done step by step process. First, uploading of images to the application is done. After uploading the image files, it reads the all images and converts into features and then splits those features to generate train and test model. After training and testing, we are applying EML algorithm on train image to generate model. EML trained model will be applied on test images to calculate prediction accuracy. Later application of adaptive safe semi supervised algorithm on train image to generate model is done. It calculates the accuracy of semi supervised algorithm and now we will compare the accuracy of both algorithms and generates a bar graph.

ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

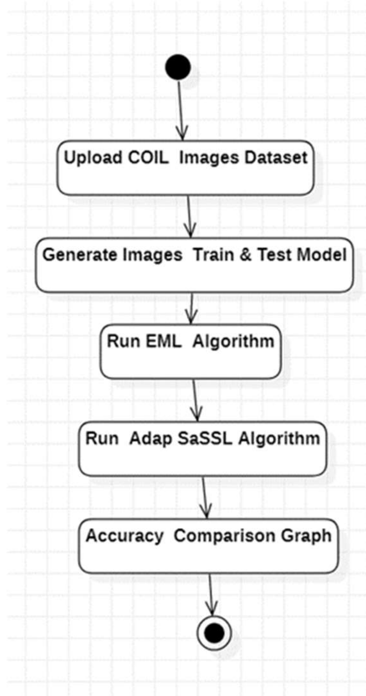
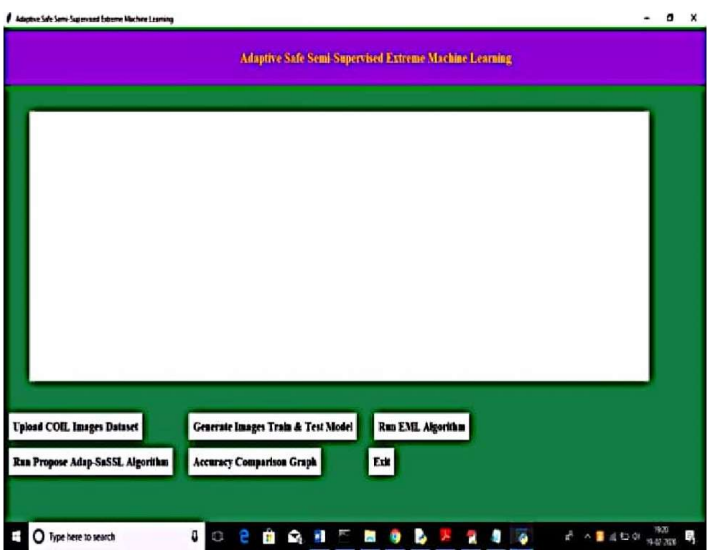



Fig: flow diagram of implementation

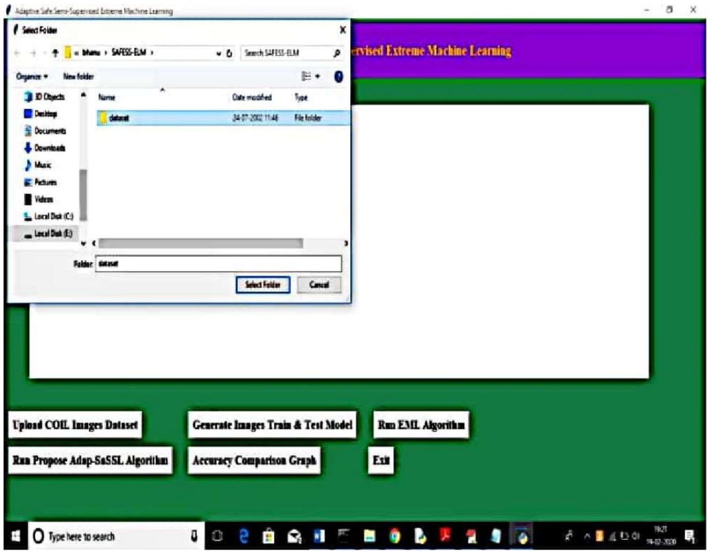
Here we used total 1440 images dataset, which are from 20 different classes. In 1440 images 1152 images are for training and 288 for testing.

Step 1: In above screen click on 'Upload COIL Images Dataset' button to upload images folder

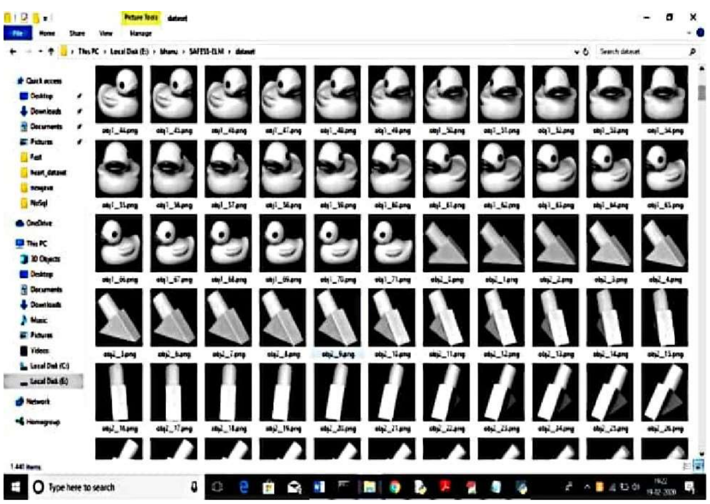



Step 2: In above screen I am uploading 'dataset' folder which contains 1500 COIL images belongs to 20 different classes.

ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

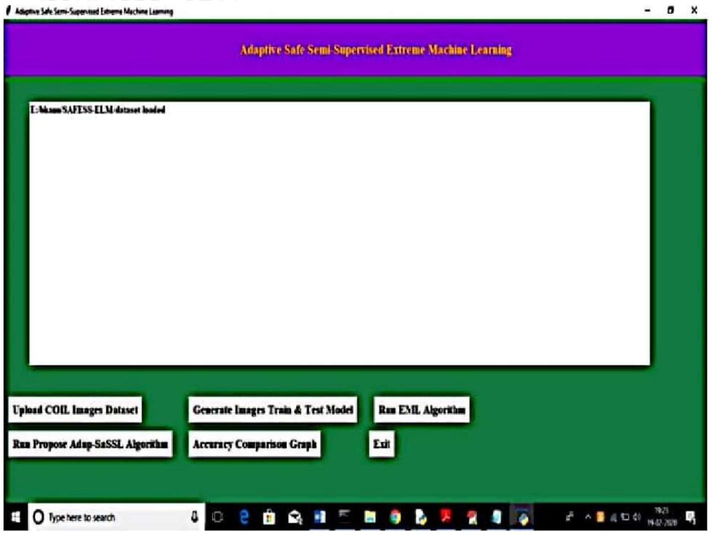



Step 3: In above dataset folder screen we can see all images loaded to application. In image name we can see image class and image name separated by underscore symbol. Obj 1 is the class name and 55 is the name of image which means 55 image belongs to class Obj 1.

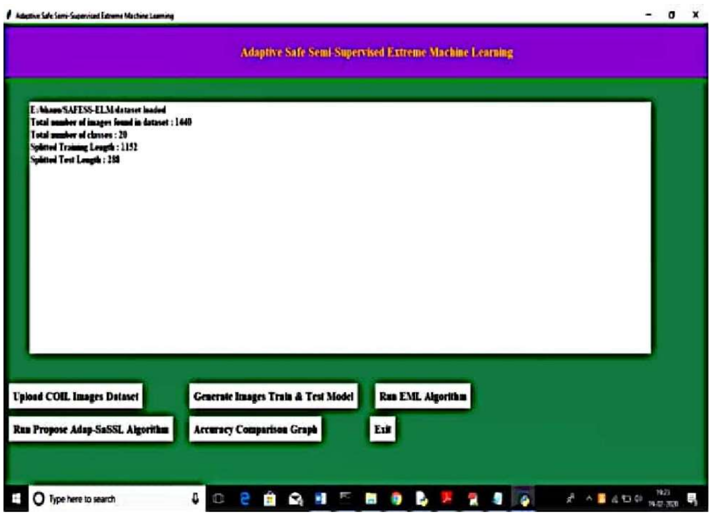


Step 4: Now click on 'Generate Images Train & Test Model' button to generate train and test model on image.

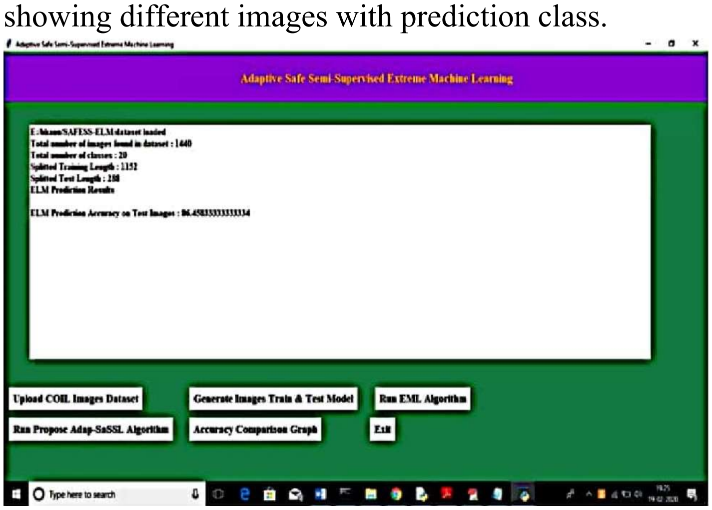
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693




Step 5: In above screen we can see total 1440 images are there from 20 different classes. Application using 1152 images for training and 288 for testing, Now click on 'Run EML Algorithm' button to calculate EML prediction accuracy on test images.

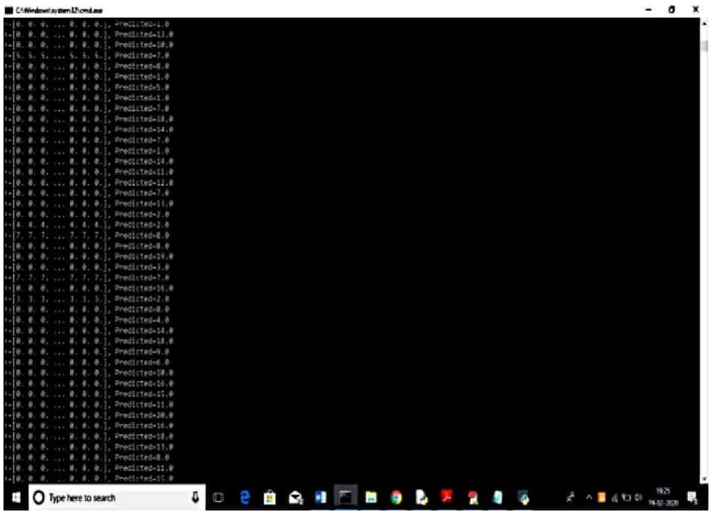



Step 6: In above screen we can see from ELM we got 86% accuracy and below screen showing different images with prediction class.

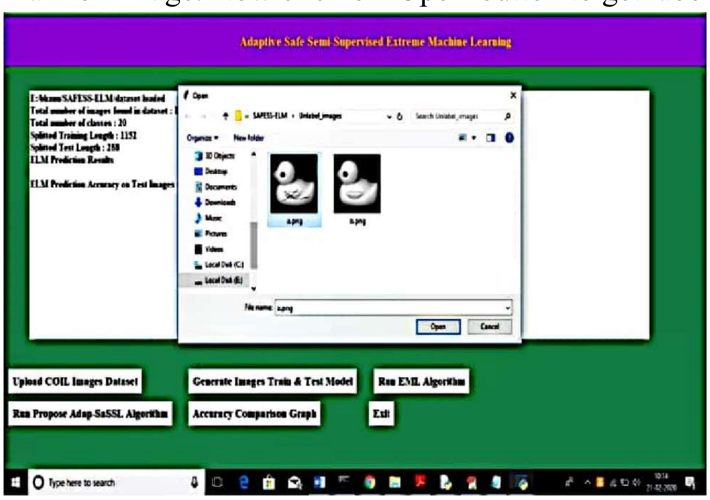


SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

Step 7: In above screen different X test images predicted in different classes. Now click on 'Run Propose Adap-SaSSL Algorithm' button to run adaptive technique.



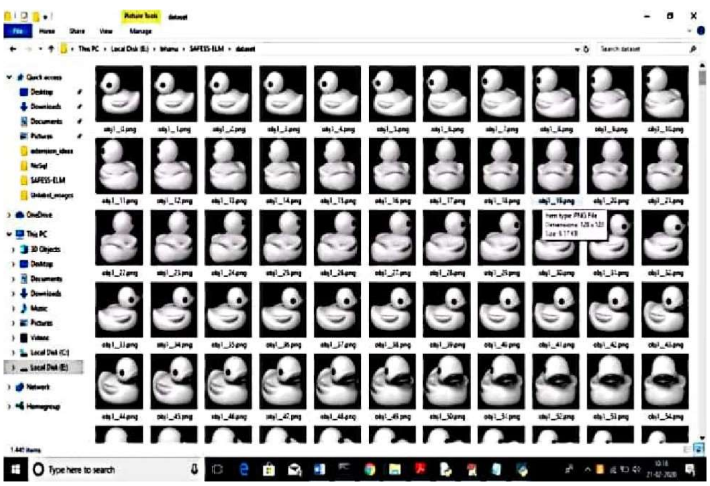
Step 8: In above screen I am uploading one Image which has only name as 'a.png' but there is no class name assign to it and to confuse classifier or to check it performance I put some mark on image. Now click on 'Open' button to get label for that image.



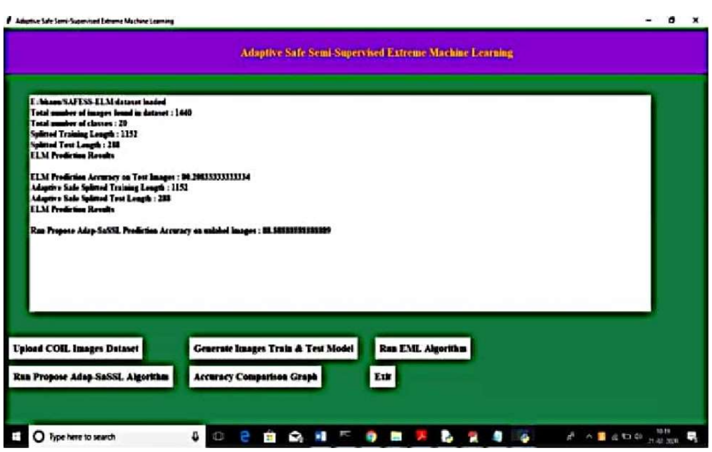

Step 9: In above screen we can see to uploaded image it assign label as 1.0 and in below dataset images we can see all ducks belongs to class 1.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693


Step 10: In above screen all duck image are in under Obj 1 class and propose technique assigning correct label to unlabeled image. See below screen for both algorithm accuracy.



Step 11: In above screen we got ELM accuracy as 80% and propose Adaptive SaSSl got 88% accuracy. This accuracy will vary from time to time as classifier will take 80% of data as training and 20% of data for testing. Classifier will choose 20% test randomly so accuracy will vary every time. Now click on 'Accuracy Comparison Graph' button to get below graph.

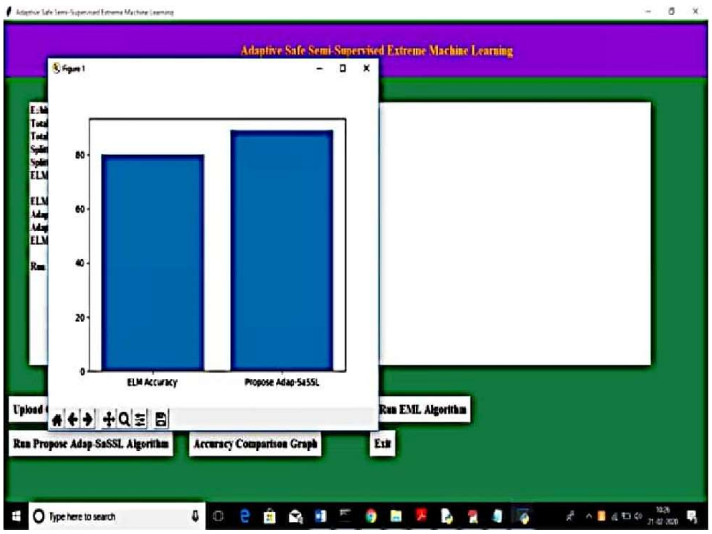


DOI: 10.5281/zenodo.6877693

**Result**

The proposed method classification shows better ELM, SSELM, SAFE-SSELM on all datasets. When we compare our method to ELM our method is effectively used for unlabeled samples for better performance. Further it was found that our method has similar performance when compared to other semi-supervised algorithms.

Here it utilizes the adaptive graphing mechanism and safety control strategy used to improve the accuracy compared to traditional methods. It was observed that safe semi supervised extreme machine learning has better classification and accuracy than extreme machine learning and semi supervised extreme machine learning. In the above graph x-axis represents algorithm name and y-axis represents accuracy of those algorithm.




In the above graph we can observe the graph between ELM and semi supervised algorithms accuracy comparision. By observing the graph ,we can observe that accuracy is more in semi supervised algorithm. Here we used labeled and unlabeled images so ELM is mostly used for labeled data whereas, SSL is used for both labeled and unlabeled data.

## 4. Conclusion

In this paper, the proposal of safe semi-supervised learning framework is done ,the basic idea of these frameworks is to calculate the manifold graphs of learning and the safety degree of each and every unlabeled sample. In this we have proposed safe semi supervised framework through machine learning. The results of the experiment on different datasets shows the performance of safe-semi supervised framework through machine learning is never importantly inferior to extreme machine learning. In future work, we will discuss other methods to analyze the trouble of unlabeled samples.

## References

[1] B. Geng, D. Tao, C. xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 6, pp. 1227_1233, Jun. 2012.

SEYBOLD Report
ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

[2] D. Yu, B. Varadarajan, L. Deng, and A. Acero, "Active learning and semisupervised learning for speech recognition: A uni_ed framework using the global entropy reduction maximization criterion," Comput. Speech, Lang., vol. 24, no. 3, pp. 433_444, 2010.

[3] F. Roli and G. L. Marcialis, SSSemi-supervised PCA-based face recognition using self-training," in Proc. Joint IAPR Int. Workshops Struct. , Syntac- tic, Stat. Pattern Recognit. (SSPR), vol. 4109, Hong Kong, Aug. 2006, pp. 560_568.

[4] H. Gan, N. Sang, and R. Huang, "Self-training-based face recognition using semi-supervised linear discriminant analysis and af_nity propagation," J. Opt. Soc. Amer. A, Opt. Image Sci., vol. 31, no. 1, pp. 1_6, 2014.

[5] W. Tsang and J. T. Kwok, &SLarge-scale sparsi_ed manifold regularization," in Proc. Adv. Neural Inf. Process. syst., 2007, pp. 1401_1408.

[6] K. Zhang, J. T. Kwok, and B. Parvin, "Prototype vector machine for large scale semi-supervised learning," in Proc. Int. Conf. Mach. Learn., 2009, pp. 1233_1240.

[7] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," .L Mach. Learn. Res., vol. 7, pp. 2399_2434, Nov. 2006.

[8] O. Chapelle, B. Schlkopf, and A. Zien, "Semi-supervised learning," in Handbook on Neural Information Processing. Berlin, Germany: Springer, 2013.

[9] P. K. Mallapragada, R. Jin, A. K. Jain, and Y. Liu, SemiBoost: Boosting for semi-supervised learning," IEEE Trans. Pattern Anal. Mach. Untell., vol. 31, no. 11, pp. 2000 2014, Nov. 2009.

[10] S. Van Vaerenbergh, 1. Santamaria, and P. E. Barbano, ``Semi-supervisedhandwritten digit recognition using very few labeled data," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., vol. 7882, May 2011, pp. 2136_2139.

[II] X. Zhu, Semi-supervised learning literature survey," Compute Sci., Univ. Wisconsin-Madison, Tech. Rep., 2008.

[12] Y. cao, H. He, and H. Huang, "LIFT: A new framework of learning from testing data for face recognition," Neurocomputing, vol. 74, no. 6, pp. 916_929, 2011.

[13] Y.-F, Li and Z-H. Zhou, "Improving semi-supervised support vector machines through unlabeled instances selection," in Proc. AAAI Conf. Artif. Intell., AAAI Press, 2011, pp. 386_391.

ISSN: 1533 - 9211
DOI: 10.5281/zenodo.6877693

[14] Y.-F. Li and Z.-H. Zhou, AsTowards making unlabeled data never hurt," in Proc. Int. Conf. Mach. Learn, Omnipress, vol. 37, 2011, pp. 1081_1088.

[15] Y. Wang, Y. Meng, Y. Li, S. Chen, Z. Fu, and H. xue, s Semi-supervised manifold regularization with adaptive graph construction," Pattern Recog- nit. Lett., vol. 98, pp. 90_95, Oct. 2017.