Alzheimer Disease diagnosis using Machine Learning Strategies

G. Stalin Babu
Asst. Professor
Aditya Institute of
Technology &
Management
Tekkali

Srujana Kurasakutla
Student
Aditya Institute of
Technology &
Management,
Tekkali

Jagadeesh Palli
Student
Aditya Institute of
Technology &
Management,
Tekkali

Maddi Gnana
Deepika
Student
Aditya Institute of
Technology &
Management,
Tekkali

ABSTRACT

Alzheimer Disease which is also called as Senile Dementia. In a sentence, we can conclude that the ability of an individual to function independently is a continuous deterioration in thought, behavioral and mental abilities. In our paper, Brain Imaging via magnetic resonance imaging (MRI), is used for evaluation of patients with suspected AD. We have used Tadpole Dataset which contains information of both AD and non-AD patients. Some studies have suggested that characteristics of MRI may predict the rate of decline in AD and may guide future therapy. However, clinicians and researchers will need to use machine learning techniques that can accurately predict a patient's progress from mild cognitive impairment to dementia in order to reach that stage. The outcome of this paper will help us to detect the disease in earlier stages by finding the accuracy of machine learning algorithms and determining the attribute that helped us in giving a maximum accuracy rate.

Keywords
Machine Learning, Alzheimer Disease

1. INTRODUCTION

Alzheimer's disease has become a widespread neurodegenerative brain disease among older people in recent years. There are around 44 million dementia patients worldwide, according to a report published by Alzheimer's Disease International, and by 2030 the figure will reach 76 million and by 2050, 135 million. Among these patients, Alzheimer's disease (AD) patients account for 50 percent to 75 percent [1] characterized by insidious onset and progressive impairment of episodic memory [2]. Mild cognitive impairment (MCI) is a disorder in which a person has slight but significant changes in the ability to think. Individuals with MCI have a greater chance of developing AD than those without it [3].

Although there are no medicines to cure AD, certain medicines have been used to delay the onset of certain symptoms and minimize patients ' psychological impact, such as memory loss [4]. Precise diagnosis of patients with AD or MCI is therefore very important in the early stage.

In Present scenario, Machine learning and pattern recognition approaches have currently been widely used in the development of a computer-aided brain disease diagnostic device with neuroimages such as magnetic resonance imaging (MRI)[5], positron emission tomography (PET)[6], functional MRI (fMRI)[7], and diffusion tensor imaging (DTI)[8].Studies have shown that structural MRI in clinical practice is the most comprehensive modality of imaging [9] and it is also useful to monitor different phases of AD[10]. Our method is therefore assessed on the structural MR images.

2. HISTORY

Alios Alzheimer, a trained physician and neuroanatomist, discovered this Alzheimer's disease on November 3, 1906. This disease is recorded as " A peculiar severe disease process of the cerebral cortex ". He identified a 50-year-old Auguste D, a woman he had followed from her admission for anxiety, progressive sleep, memory loss and violence until her death five years later. His report noted distinctive plaques and neurofibrillary tangles, as well as an autopsy that nerve cells in and around the brain were shrinking.

In 1931, scientists allowed brain cells to be examined in greater detail using magnification electron microscope up to 1 million times. Cognitive assessment scales were developed in 1968 to assess the disability and estimate the volume of damaged brain tissues. In 1983, the NIA (National Institute for Aging) developed an awareness program on Alzheimer's disease and established national wide disease centers and Alzheimer's research network.

In 1993, FDA (Food & Drug Administration) approved the first drug called Cognex. This drug treats the effects of memory loss and dementia [20].

3. SYMPTOMS
There are five stages in the Alzheimer disease which helps to understand that the person is in which stage by its symptoms.

3.1 Preclinical Alzheimer's disease:
That starts long before any symptoms appear. It is commonly found in research settings. During this stage we can not notice the symptoms. This stage can last years or may last for decades. New imaging techniques can now recognize the deposits of a protein called amyloid-beta, which is a characteristic of Alzheimer's disease. Biomarkers can assess the increased risk of diseases and can usually support Alzheimer's disease diagnosis. And the genetic tests also became more important as new treatment and more beneficial for people suffering from Alzheimer's disease with this initial stage.

3.2 Mild cognitive impairment (MCI):
People with mild cognitive impairment have minor changes in their ability to remember and to think. Patients with MCI have lapses in memory when it comes to information that is usually easily remembered. And they also have difficulty deciding

how much time a job takes. Ability to make sound decisions can get harder.

3.3 Mild dementia

In this moderate dementia, people may experience memory loss from recent events, difficulty during problem-solving, complex tasks and sound decisions, personality changes, difficulty in organizing and communicating the thoughts and getting lost or screwing up the items.

3.4 Moderate dementia

At this point, people may experience even greater memory loss, need to help with certain daily activities, undergo major personality and behavioral changes

3.5 Severe dementia

This is the last and severe stage of Alzheimer's disease. In this stage, mental function continuous to decline and the disease has growing impact on movement and physical capabilities [21].
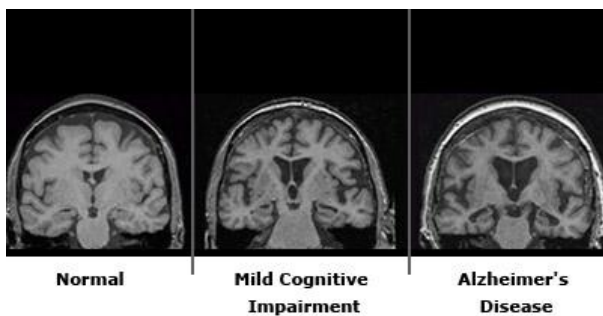

Figure 3 : Normal vs MCI comparison

4. PRESENT SCENARIO OF ALZHEIMER'S DISEASE

The number of people living with Alzheimer's disease is growing rapidly and only about one in four people get diagnosed with the disease. About 44 million people worldwide are projected to be living with this disease. An estimated 5.5 million people of all ages in the U.S. suffer from Alzheimer's disease. Approximately 5.3 million people are 65 and older and 200,000 are young and have Alzheimer's disease early onset.

There is larger in population and larger in impact of illness. At present, U.S., develops Alzheimer's illness each sixty six seconds. it's thought that by the center of the century, somebody within the U.S., can develop the illness each thirty three seconds and also the total range of individuals with Alzheimer's illness within the U.S. rise to as high as sixteen million individuals by 2050. And in Asian nation around four million individuals were laid low with this disease[22].

5. MACHINE LEARNING

Machine learning is a man-made intelligence (AI) technology that has systems with the power to be told and improve mechanically from expertise while not express programming. Machine learning focuses on making pc programs which will access and use data for learning on their own [11].

Machine learning is an area of computing science where the machine predicts the next operation to be carried out by analyzing the data it receives. The information accessed by the software may be in the form of digitized training sets, or through communication with the environment. Machine learning algorithms is made in such the way that information is learned and foreseen as opposition static programming algorithms that need direct human instruction. totally different supervised and unsupervised techniques are recommended to unravel problems like rule-based techniques, logic-based techniques, instance-based techniques, stochastic techniques [12].

For the training of the machine learning algorithm a training data set is used to create a model. When new input data is introduced by the ML algorithm, this allows a model-based prediction. The prediction is tested for accuracy, and if the accuracy is appropriate, the Machine Learning Algorithm is applied.If the accuracy is not appropriate, the Machine Learning Algorithm will be trained over and over again with an increased training data collection [13].

5.1 Types of Machine Learning

5.1.1 Supervised Learning
This algorithm consists of a target / outcome factor (or dependent variable) to be computed from a set of predictors (independent variables).We construct a function using these set of variables to map inputs to desired outputs. The training process continues until a desired degree of reliability on the training data is obtained by the prototype [14].

5.1.2 Unsupervised Learning
We have no goal or outcome factor to predict / estimate in this algorithm.It is used to clustering population in different groups, which is commonly used to segment customers in different groups for direct intervention[14].

5.1.3 Reinforcement Learning
Reinforcement learning describes a category of issues wherever associate agent operates in associate environment and should learn to control using feedback.

Reinforcement Learning to improve is training what to do—how to map situations to behaviour — so as to optimize a signal of numerical rewards. The learner isn't told that actions to require, however instead should discover that actions yield the foremost reward by making an attempt them[24].
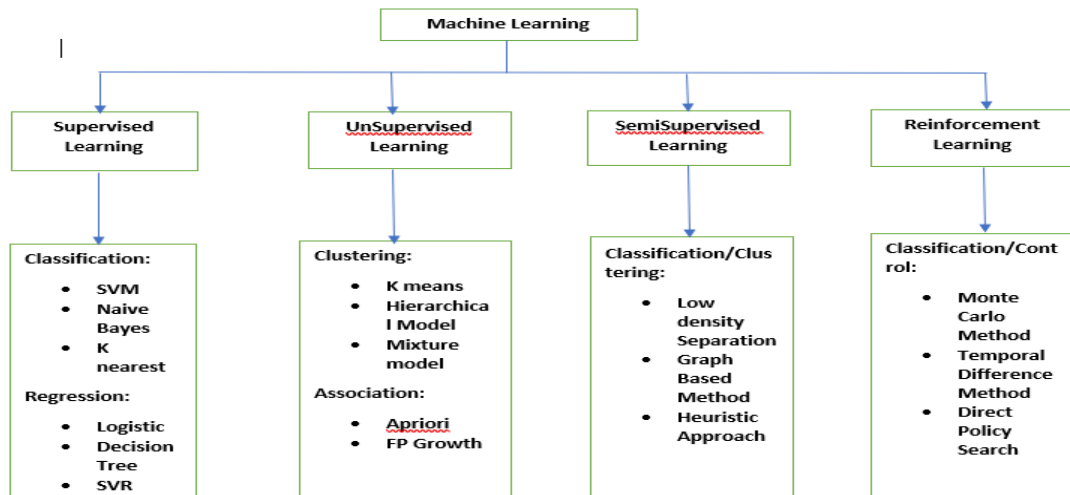
Figure 5.1 :Types of Machine Learning

5.2 Support Vector Machine (SVM):

Support vector machine (SVM) could be a supervised machine learning algorithmic rule. SVM is used for each classification or Regression issues[23].

5.2.1 SVM Classification

It's a type of classification during this rule, the data item is premeditated as some extent in n-dimensional space (where n is that the range of options you have) with every feature being the worth of a specific coordinate.

For example, if we tend to had solely 2 options like a person's height and hair length, we might 1st plot these 2 variables in 2-dimensional area wherever every purpose has two coordinates (these are referred to as Support Vectors).
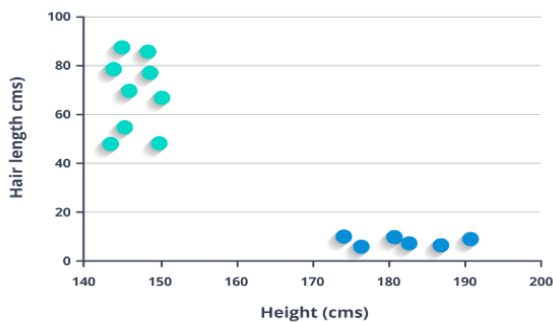

Figure 5.2.1.1: Data items plotted in n-dimensional space

Now we're aiming to realize some line separating the {data| the info| the information} between the 2 data classes that are severally graded. this can be a line in order that the distances in every of the 2 classes from the nearest point are as distant as attainable [14].
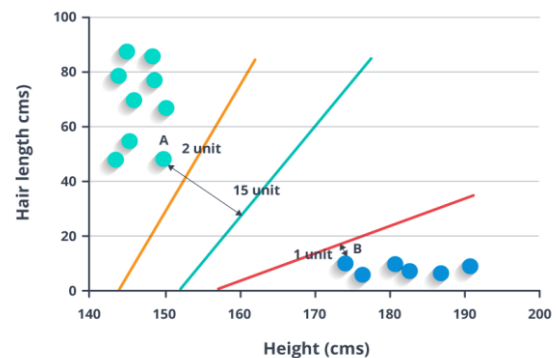

Figure 5.2.1.2 : Classification using Hyperplanes

5.2.2 SVM Regression

If we tend to get a selected training set{ $(x1,y1),....,(xi, yi)$ } , say X x R, here X -> space of input patterns. Our objective in SV regression is to search out a fitting function f(x), having a less deviation to the relevant training data set than the target (xi) acquired. The function ought to be fairly flat. Or any error may be said to be less than perfect. The linear function (f)-f (x) = (w, x) + b with, w X, b R Here (·, ·) is that the dot product of X, flatness is outlined by w during this case. to make sure that we've got to stay the norm to a minimum [12].

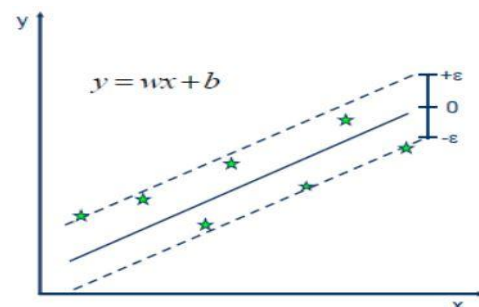**Figure : 5.2.2.1 Support Vector Regression**

5.2.3 SVM Kernel Functions

SVM algorithms use a series of kernel defined mathematical functions. The kernel's function is to take input data and convert it into the form necessary.Specific kernel function types are used by specific SVM algorithms. These can be

different types of functions. For eg, linear, polynomial, RBF function, and sigmoid function. The functions of the kernel return the inner product in an appropriate feature space between two points. Therefore, by establishing a notion of similarity, with low computational costs even in very large spaces[15].

Linear:

$$K(Xi, Xj) = Xi \cdot Xj$$

Polynomial:

$$K(Xi, Xj) = (\gamma Xi \cdot Xj + C)^d$$

RBF:

$$K(Xi, Xj) = \exp(-\gamma |Xi - Xj|^2)$$

Sigmoid:

$$K(Xi, Xj) = \tanh(\gamma Xi \cdot Xj + C)$$

Where $K(Xi, Xj) = \phi(Xi) \cdot \phi(Xj)$

The kernel function denotes the dot product of input data points mapped by transformation into the higher dimensional feature space ϕ.Gamma is an adjustable parameter for certain functions of the kernel[16].

5.3 Decision Tree

Decision tree constructs models of regression or classification as a tree structure. It breaks down a dataset into smaller and smaller sub-sets whereas at identical time progressively developing an associated decision tree. the end result's a tree with nodes of decision and nodes of leaf. There ar 2 or additional branches in a decision node. Leaf node reflects a numerical target decision. in a tree that corresponds to the most effective predictor referred to as root node, the very best decision node. each categorical and numerical information will be handled by decision trees [17].

5.3.1 Decision Tree Classification

The ID3 (Iterative Dichotomiser 3) is that the basic formula accustomed build the decision tree. To construct the decision tree, the ID3 formula uses Entropy and information Gain.

Steps:

i. Calculate the target's entropy.

ii. Then the dataset is split on the varied attributes. The entropy is set for every branch. Then, to induce

total entropy for the split it's applied proportionately. Before the split, the ensuing entropy is

separated from the entropy. The result is information gain or decreasing entropy.Gain(T,X)=Entropy(T)-Entropy(T-X)

iii. select the attribute with the best information gain because the decision node, divide the dataset by its branches and perform a similar procedure for every branch.

iv. A zero entropy branch could be a leaf node. A branch with over zero entropy must be any split.

v. For non-leaf branches, the ID3 formula runs recursively till all information is assessed [18].

5.3.2 Decision tree regression

The ID3 (Iterative Dichotomiser 3) is that the basic algorithmic program accustomed build the decision tree. To construct the decision tree, the ID3 algorithmic program uses standard deviation Reduction (SDR).

Steps:

i. first we tend to calculate the target's standard deviation.

ii. we tend to split the data sets on varied attributes once this. for every branch the resulting standard deviation is deducted from the standard deviation before splitting. That's SDR.

SDR(T,X) = S(T) — S(T, X)

iii. it's vital to choose the attribute with the most important SDR as the decision node.

iv. Dataset should be divided according to the values of the attributes elect. If the standard deviation is bigger than zero, we tend to more split a branch set.

v. The method continues to run in recursion till all the data are processed[12].

6. TADPOLE DATASET

Alzheimer's Disease Neuroimaging Initiative (ADNI) give datasets that can be utilized for different Alzheimer's Disease related examinations. The dataset contains a lot of information that was produced specifically for the TADPOLE Grand Challenge. The dataset consists of various features that can be classified as: demographic, cognitive tests, and biomarkers obtained from MRI and PET scans. The biomarkers are numerical values of the brain volumes identified using the FreeSurfer software.

The dataset consists of approximately 1900 features and we only use 245 of these features to train the model. Specifically, we use all the demographic features, three cognitive features, and all the features from MRI biomarkers. The data are categorized as normal (NL), moderate cognitive impairment (MCI), and dementia.[19]

6.1 Dataset Description

Table 1 : Tadpole Dataset

Dataset label	Dementia	MCI	NL	Total Sample size
Completed Data	1454	3298	2140	7359
After Processing	104	324	254	682
MCI to NL	104	0	578	682
MCI to Dementia	428	0	254	682

7. DATA PRE-PROCESSING

A pre-processing of the data was required due to the amount of missing feature values in the original dataset. Firstly, any sample that has missing values is removed. Also, the features that have text labels are converted to numerical labels. After removing these missing values, there are some very few (less than 10) sample points with additional labels. These are the cases where, for example, the label was changed from "MCI to Dementia" or "MCI to NL" in the dataset. The labels were translated to the appropriate label in such cases, i.e. the sample was relabeled as Dementia in the case of "MCI to Dementia." As these are very few situations, this change does not reflect major changes in the dataset. Therefore, since the

goal is to identify only Alzheimer's disease, the MCI and NL labels are classified as non-demented or regular, resulting in a data set of only two groups. One choice was to mark a patient as having or not having a cognitive impairment. In this case, MCI and Dementia labels have been treated as one class while NL has been treated as another. Both were attempted but the main goal was to identify Alzheimer's disease.

There is a high variance in 245 features with 682 sample points. Therefore, two methods of compression were attempted, namely PCA and LDA. Both methods were used to minimize the number of characteristics to 10. PCA recognizes the main components that have the most variance and are not dependent on labels. LDA is going to find features that optimize the distinction between different classes and is therefore dependent on the label.

7.1 Data Splitting

The data set was divided into a train set (70 percent of the samples) and a test set (30 percent of the sample) to train different models and determine the best model. To obtain new train and dev sets, we applied k-fold cross-validation on this training set, with k = 5.

8. RESULTS

To obtain new train and dev sets, we applied k-fold cross-validation on this training set, with k = 5. When comparing the accuracy levels between Support Vector Machine (SVM) and Decision Tree algorithms The Decision Tree outperformed well achieving 0.8 test accuracy when compared to the SVM with 0.75 test accuracy. When testing SVM accuracy levels, LDA was shown to be useful in reducing the model's variance by reducing the dimension from 245 to 10 features and to some extent reduce the test accuracy. In this situation, PCA has also developed the standard, though by a smaller percentages. If we had more time, we would compare more number of classifications with various algorithms and also calculate the respective accuracy levels. Given below are the confusion matrices for SVM and Decision Tree respectively.
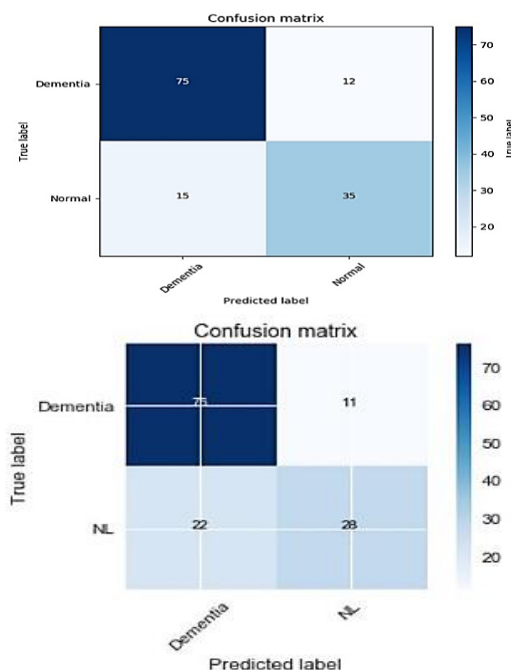

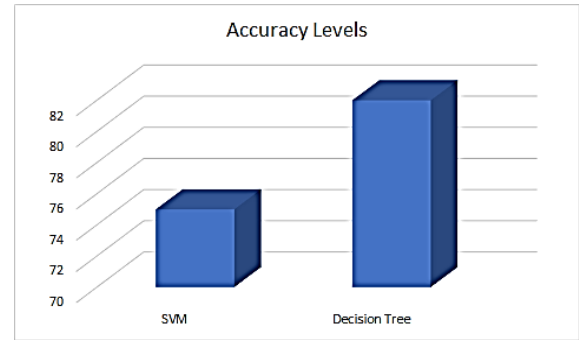**Figure 8.1 : Confusion Matrix of Decision Tree and SVM**

Figure 8 : Accuracy Levels of SVM and Decision Tree

9. CONCLUSION

The outcomes of this project will help us to detect the disease in earlier stages by finding the accuracy of machine learning algorithms and determining the attribute that helped us in giving a maximum accuracy rate. Some studies have suggested that characteristics of MRI may predict the rate of decline in AD and may guide future therapy. However, clinicians and researchers will need to use machine learning techniques that can accurately predict a patient's progress from mild cognitive impairment to dementia in order to reach that stage.

10. REFERENCES

[1] M. Prince, E. Albanese, and M. Guerchet, World Alzheimer report, 2014.

[2] K. Blennow, "Alzheimer's disease," Lancet, vol. 368, no. 9533, pp. 387–403, 2006.

[3] Alzheimer's association, "2014 Alzheimer's disease facts and figures," Alzheimer's and Dementia, vol. 10, no. 2, 2014.

[4] K. G. Yiannopoulou and S. G. Papageorgiou, "Current and future treatments for Alzheimer's disease," Therapeutic Advances in Neurological Disorders, vol. 6, no. 1, pp. 19–33, 2013.

[5] D. Zhang and D. Shen, "Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers," PLoS ONE, vol. 7, no. 3, Article ID e33182, 2012.

[6] A. Nordberg, J. O. Rinne, A. Kadir, and B. Långström, "The use of PET in Alzheimer disease," Nature Reviews Neurology, vol. 6, no. 2, pp. 78–87, 2010.

[7] H.-I. Suk, C.-Y. Wee, and D. Shen, "Discriminative group sparse representation for mild cognitive impairment classification," in Machine Learning in Medical Imaging, pp. 131–138, Springer, 2013.

[8] S. Keihaninejad, H. Zhang, N. S. Ryan et al., "An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to Alzheimer's diseas," NeuroImage, vol. 72, pp. 153–163, 2013.

[9] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain MRI," Medical Image Analysis, vol. 18, no. 5, pp. 808–818, 2014.

[10] M. Goryawala, Q. Zhou, W. Barker, D. A. Loewenstein, R. Duara, and M. Adjouadi, "Inclusion of neuropsychological scores in atrophy models improves

diagnostic classification of Alzheimer's disease and mild cognitive impairment," Computational Intelligence and Neuroscience, vol. 55, Article ID 865265, pp. 856–867, 2015.

[11] https://expertsystem.com/machinelearningdefinition/#targetText=Machine%20learning%20is%20an%20application,use%20it%20learn%20for%20themselves.

[12] "https://ieeexplore.ieee.org/document/8320256"

[13] "https://www.edureka.co/blog/what-is-machine-learning/"

[14] "https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/"

[15] "https://data-flair.training/blogs/svm-kernel-functions/"

[16] "http://www.statsoft.com/textbook/support-vector-machines

[17] "https://www.saedsayad.com/decision_tree.htm"

[18] https://medium.com/@rishabhjain_22692/decisiontrees-it-begins-here-93ff54ef134

[19] "https:/Ida.loni.usc.edu/login.jsp?project=ADNI&page=HOME&logout=true"

[20] "http://www.alzheimers.net/history-of-alzheimers/"

[21] https://www.mayoclinic.org/diseasesconditions/alzheimers-disease/in-depth/alzheimers-stages/art-20048448

[22] "PresentScenario of Alzheimer's"[Online] Available:https://www.who.int/medicines/areas/priority_medicines/BP6_11Alzheimer.pdf

[23] G Stalin Babu , S N Tirumala Rao, R Rajeswara Rao."Exploring of Classification Methods for Early Detection of Alzheimers Disease"International Journal of Engineering and Advanced Technology (IJEAT)Vol:8Issue:6 pgno:5206-5211.Blue Eyes Intelligence Engineering & Sciences Publication(2019)

[24] https://machinelearningmastery.com/types-of-learning-in-machine-learning