

HUMAN ANNOTATOR FOR IMBALANCED DOSSIER

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree
in
Computer Science and Engineering
By

SAMPATH (38110478)

SSAHUKARA(38110156)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade “A” by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

CHENNAI - 600 119

APRIL – 2020



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with -All grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of

, **SAMPATH (38110478), SSAHUKARA(38110156** who

carried out the project entitled **“HUMAN ANNOTATOR FOR IMBALANCED DOSSIER”** under my supervision from **NOVEMBER 2019** to **APRIL 2020**.

Internal Guide

Dr. Damodaran

Head of the Department

Dr. S. VIGNESWARI, M.E., Ph.D.,

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I SSAHUKARA(38110156)

hereby declare that the Project Report entitled **“HUMAN ANNOTATOR FOR IMBALANCED DOSSIER”** done by me under the guidance of **Dr. Damodaran** is submitted in partial fulfillment of the requirements for the award of **Bachelor of Engineering degree in Computer Science and Engineering.**

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. SASIKALA M.E., Ph.D., Dean**, School of Computing, also **Dr. S. VIGNESHWARI M.E., Ph.D.,** and **Dr. L. LAKSHMANAN M.E., Ph.D., Heads** of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. Damodaran** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

An imbalanced dataset is relevant primarily in the context of supervised machine learning involving two or more classes. Imbalance means that the number of data points available for a different class is different, Imbalanced data sets is a special case for classification problem where the class distribution is not uniform among the classes. Imbalanced classes are a common problem in machine learning classification where there is a disproportionate ratio of observations in each class. Class imbalance can be found in many different areas including medical diagnosis, spam filtering, and fraud detection. Typically, they are composed of two classes, the majority (negative) class, and the minority (positive) class. Information about experimental studies sets. These types of data sets are typically found on websites that collect and aggregate data sets. These aggregators tend to have data sets from multiple sources, without much creation. In this case, that's a good thing too much creation gives us overly neat data sets that are hard to label. Active learning is undoubtedly effective, but several recent studies have indicated that active learning is failed when it is applied to data. In our project, Human Annotator will collect the data's from the public post and he will separate labeled and unlabelled data sets. User needs to register their details and they can view their learning materials. The labelled and unlabelled data's have been analyzed by human annotator and matched with labelled appropriate data sets are achieved and it will be available for learners.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	v
	LIST OF FIGURES	ix
	LIST OF TABLES	x
	LIST OF ABBREVIATIONS	xi
1	INTRODUCTION	1
	1.1 OUTLINE OF THE PROJECT	1
	1.2 PURPOSE OF THE PROJECT	1
	1.3 ACTIVE LEARNING PARADIGM	2
	1.4 MACHINE LEARNING	4
	1.4.1 Machine Learning Methods	5
	1.4.2 Applications of Machine Learning	5
	1.4.3 Advantages	5
	1.5 SYSTEM REQUIREMENTS	6
	1.5.1 Hardware Requirements	6
	1.5.2 Software Requirements	6
2	LITERACTURE SURVEY	7
	2.1 SURVEY 1	7
	2.2 SURVEY 2	7
	2.3 SURVEY 3	8
	2.4 SURVEY 4	9
	2.5 SURVEY 5	10
	2.6 SURVEY 6	10
3	METHODOLOGY	12
	3.1 SYSTEM ANALYSIS	12
	3.1.1 Analysis Model	12

3.2 EXISTING SYSTEM	14
3.2.1 Disadvantage	14
3.3 PROPOSED SYSTEM	15
3.3.1 Advantage	15
3.4 MODULES OF THE PROJECT	16
3.4.1 Authentication and Authorization	16
3.4.2 Material Upload	16
3.4.3 Active Learning	17
3.4.4 Active Learning with Extreme Learning Machine	17
3.4.5 Learning Material	17
3.5 FEASIBILITY REPORT	17
3.5.1 Technical Feasibility	18
3.5.2 Operational Feasibility	19
3.5.3 Economic Feasibility	19
3.6 SOFTWARE REQUIREMENT SPECIFICATION	19
3.6.1 Developers Responsibilities Overview	20
3.6.2 Functional Requirements	20
3.6.3 Non-Functional Requirements	20
3.6.4 Performance Requirements	21
3.7 THE .NET FRAMEWORK ARCHITECTURE	21
3.7.1 Common Language Runtime Engine	22
3.7.2 Language Independence	23
3.7.3 Framework Class Library	23
3.7.4 Simplified Deployment	23
3.7.5 Security	23
3.7.6 Portability	23
3.7.7 Common Language Specification	24
3.8 SQL SERVER 2014	24
3.9 SYSTEM DESIGN	25
3.10 SYSTEM TESTING AND IMPLEMENTATION	26

	3.10.1 Strategic Approach to Software Testing	26
	3.11 SYSTEM SECURITY	27
	3.11.1 Security in Software	27
	3.11.1.1 Client Side Validation	28
	3.11.1.2 Server Side Validation	28
4	RESULTS AND DISCUSSIONS	29
	4.1 NORMILIZATION	29
	4.2 E-R DIAGRAMS	32
	4.3 DATA FLOW DIAGRAM	37
	4.4 ALGORITHM	38
5	CONCLUSION AND FUTURE WORK	41
	REFERENCES	42
	APPENDIX	43
	A. SAMPLE CODE	43
	B. SCREENSHOTS	53
	C. PUBLICATION WITH PLAGARISM REPORT	59

LIST OF FIGURES

FIGURE	TITLE	PAGE
NO.		NO.
1.1	Overview of Machine Learning	4
3.1	Spiral Model	13
3.2	System Architecture	16
3.3	.NET Architecture	22
3.4	Server Architecture	24
4.1	ER Diagram for Human Annotator	33
4.2	Flowchart of Human Annotator	34
4.3	Sequence Diagram for Human Annotator	35
4.4	Activity Diagram for Human Annotator	36
4.5	Level-0 Data Flow Diagram	37
4.6	Level-1 Data Flow Diagram	38

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
4.1	Design for User Register	29
4.2	File Upload	31

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPANSION
ANSI	American National Standards Institute
AOW-ELM	Active Online Weighted-Extreme Learning Machine
ASP	Active Server Page
ASR	Automatic Speech Recognition
BCL	Base Class Library
BLOBs	Binary Large Objects
CBIR	Content Based Image Retrieval
CLR	Common Language Runtime
ELM	Extreme Learning Machine
FCL	Framework Class Library
MLP	Multiple Level Perceptron
SQL	Structured Query language
SRS	Software Requirement Specification
SVM	Support Vector Machine
WCF	Windows Communication Foundation

CHAPTER 1

INTRODUCTION

1.1. OUTLINE OF THE PROJECT

Imbalanced data typically refers to classification tasks where the classes are not represented equally. Most of the real-world classification problems display some level of class imbalance, which happens when there are not sufficient instances of the data that correspond to either of the class labels. Therefore, it is imperative to choose the evaluation metric of your model correctly. If it is not done, then you might end up adjusting/optimizing a useless parameter. In a real business-first scenario, this may lead to a complete waste.

There are problems where a class imbalance is not just common it is bound to happen. For example, the datasets that deal with fraudulent and non-fraudulent transactions, it is very likely that the number of fraudulent transactions as compares to the number of non-fraudulent transactions will be very much less. And this is where the problem arises. When the dataset has underrepresented data, the class distribution starts skew.

Due to the inherent complex characteristics of the dataset, learning from such data requires new understandings, new approaches, new principles, and new tools to transform data. And moreover, this cannot anyway guarantee an efficient solution to your business problem.

1.2. PURPOSE OF THE PROJECT

Dealing with imbalanced datasets includes various strategies such as improving classification algorithms or balancing classes in the training data (essentially a data pre-processing step) before providing the data as input to the machine learning. The latter technique is preferred as it has a broader application and adaptation. Moreover, the time taken to enhance is often higher than to generate the required samples. But for research purposes, in our project Human Annotator will

collect the imbalanced data, he will segregate labeled and unlabelled data to provide a complete learning material.

1.3. ACTIVE LEARNING PARADIGM

Active learning is a popular machine learning paradigm and it is frequently deployed in the scenarios when large scale instances are easily collected, but labeling them is expensive and/or time-consuming. By adopting active learning, a classification model can iteratively interact with human experts to only select those most significant instances for labeling and to further promote its performance as quickly as possible. Therefore, the merits of active learning lie in decreasing both the burden of human experts and the complexity of training instances but acquiring a classification model that delivers superior or comparable performance to the model with labeling all instances.

Past research has accumulated a large number of active learning models, and generally, we have several different taxonomies to organize these models. Based on different ways of entering the unlabeled data, active learning can be divided into pool-based and stream-based models. The former previously collects and prepares all unlabeled instances, while the latter can only visit a batch of newly arrived unlabeled data at each specific time point.

According to different numbers of the labeled instances in each round, we have single-mode and batch-mode learning models. As their names indicate, the single-mode model only labels one unlabeled instance on each round, while the batch-mode labels a batch of unlabeled examples once. In addition, we have several different significance measures to rank unlabeled instances, including uncertainty, representativeness, inconsistency, variance, and error. Each significance measure has a criterion for evaluating which instances are the most important for improving the performance of the classification model. For example, uncertainty considers the most important unlabeled instance to be the nearest one to the current classification boundary; representativeness considers the unlabeled instance that can represent a new group of instances, e.g., a cluster, to be more important, and inconsistency

considers the unlabeled instance that has the most predictive divergence among multiple diverse baseline classifiers to be more significant. In addition, active learning models can also be divided into different categories according to which kind of classifier has been adopted.

Some popular classifiers, including naive Bayes, k -nearest neighbors, decision tree, multiple level perceptron (MLP), logistic regression, support vector machine (SVM), and extreme learning machine (ELM), have all been developed to satisfy the requirements of active learning. In the past decade, active learning has also been deployed in a variety of real world applications, such as video annotation, image retrieval, text classification, remote sensing, image annotation, speech recognition, network intrusion detection, and bioinformatics.

The proposed algorithm is named active online weighted ELM (AOW-ELM), and it should be applied in the pool-based batch-mode active learning scenario with an uncertainty significance measure and ELM classifier. In AOW-ELM, we first take advantage of the idea of cost-sensitive learning to select the weighted ELM (WELM) as the base learner to address the class imbalance problem existing in the procedure of active learning. Then, we adopt the AL-ELM algorithm presented in our previous paper to construct an active learning framework. Next, we deduce an efficient online learning mode of WELM in theory and design an effective weight update rule. Finally, benefiting from the idea of the margin exhaustion criterion, we present a more flexible and effective early stopping criterion. Moreover, we try to simply discuss why active learning can be disturbed by skewed instance distribution, further investigating the influence of three main distribution factors, including the class imbalance ratio, class overlapping, and small disjunction. Specifically, we suggest adopting the clustering techniques to previously select the initially labeled seed set, and thereby avoid the missed cluster effect and cold start phenomenon as much as possible. Experiments are conducted on 32 binary-class imbalanced data sets, and the results demonstrate that the proposed algorithmic framework is generally more effective and efficient than several state-of-the-art active learning algorithms that were specifically designed for the class imbalance scenario.

1.4. MACHINE LEARNING

Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computer to learn automatically without human intervention or assistance and adjust actions accordingly.

The test for a machine learning model is a validation error on new data not a theoretical test that proves a null hypothesis. Because this often uses an iterative approach to learn from data, the learning can be easily automated.

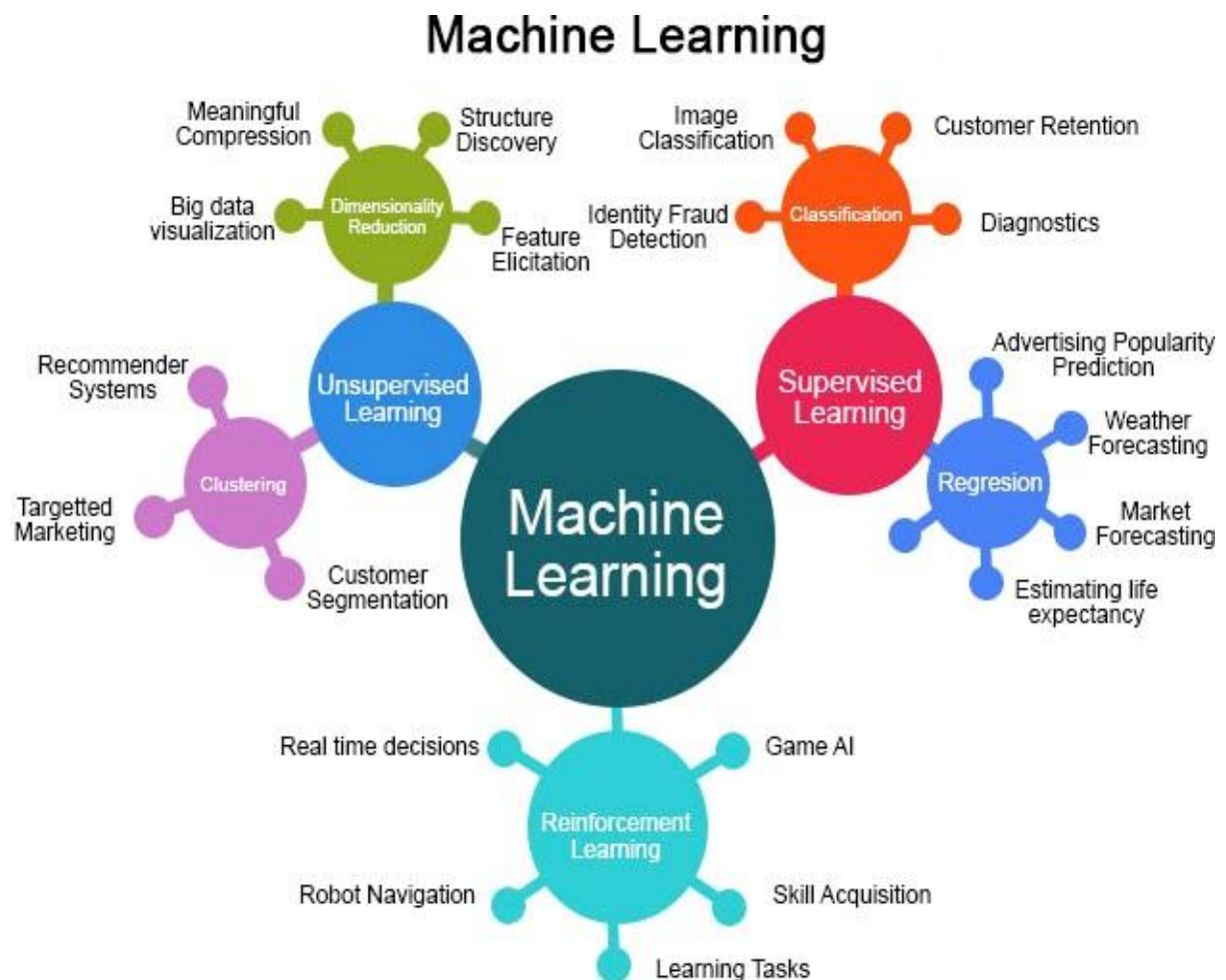


Fig 1.1 Overview of Machine Learning

1.4.1. Machine Learning Methods

Some of the methods of Machine Learning algorithm are categorized as

- **SUPERVISED LEARNING**

A Supervised learning algorithm learns from labelled training data, helps you to predict outcomes for unforeseen data. It is highly accurate and trustworthy method.

- **UNSUPERVISED LEARNING**

Unsupervised learning algorithm is the type of self - organized with the help of previously unknown patterns in dataset without pre-existing labels.

- **SEMI-SUPERVISED LEARNING**

Semi-supervised learning is the combination of both supervised and unsupervised which means labeled and unlabelled data.

- **REINFORCEMENT MACHINE LEARNING**

Reinforcement machine learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

1.4.2. Applications of Machine Learning

- Video Surveillance
- Social Media Services
- Email Spam and Malware Filtering
- Financial Services
- Health Care
- Retail
- Transportation

1.4.3. Advantages

- Computational property is cheaper and more powerful.
- Affordable data storage.
- It can analysis complex data quickly and automatically.
- It produces more accurate results.

1.5. SYSTEM REQUIREMENTS

1.5.1. *Hardware Requirements*

- RAM 4GB
- Dual Core 2.8 GHz Processor and Above
- HDD 80 GB Hard Disk Space and Above

1.5.2. *Software Requirements*

- WINDOWS OS (7 /XP)
- Visual Studio .Net 2015
- Visual Studio .Net Framework 4.5
- SQL Server 2014

CHAPTER 2

LITERATURE SURVEY

2.1. SURVEY 1

Publication: IEEE Transactions on Multimedia Visualization - based active learning for video annotation.

Authors: Hongsen Liao, Li Chen, Yibo Song, Hao Ming.

Description: Video annotation is an effective way to facilitate content-based analysis for videos. Automatic machine learning methods are commonly used to accomplish this task. Among these, active learning is one of the most effective methods, especially when the training data cost a great deal to obtain. One of the most challenging problems in active learning is the sample selection. Various sampling strategies can be used, such as uncertainty, density, and diversity, but it is difficult to strike a balance among them. In this paper, we provide a visualization-based batch mode sampling method to handle such a problem. An iso-contour-based scatterplot is used to provide intuitive clues for the representativeness and informativeness of samples and assist users in sample selection. A semisupervised metric learning method is incorporated to help generate an effective scatterplot reflecting the high-level semantic similarity for visual sample selection. Moreover, both quantitative and qualitative evaluations are provided to show that the visualization-based method can effectively enhance sample selection in active learning.

2.2. SURVEY 2

Publication: IEEE Transactions on Image Processing focuses Active learning methods for interactive image retrieval.

Authors: Philippe Henri Gosselin, Matthieu Cord.

Description: Active learning methods have been considered with increased interest in the statistical learning community. Initially developed within a classification

framework, a lot of extension are now being proposed to handle multimedia applications. This paper provides algorithms within a statistical framework to extend active learning for online content-based image retrieval (CBIR). The classification framework is presented with experiments to compare several powerful classification techniques in this information retrieval context. Focusing on interactive methods, active learning strategy is then described. The limitations of this approach for CBIR are emphasized before presenting our new active selection process RETIN. First, as any active method is sensitive to the boundary estimation between classes, the RETIN strategy carries out a boundary correction to make the retrieval process more robust. Second, the criterion of generalization error to optimize the active learning selection is modified to better represent the CBIR objective of database ranking. Third, a batch processing of images is proposed. Our strategy leads to a fast and efficient active learning scheme to retrieve sets of online images (query concept). Experiments on large databases show that the RETIN method performs well in comparison to several other active strategies.

2.3. SURVEY 3

Publication: IEEE Transactions on Geoscience Remote Sensing and An active learning approach to hyperspectral data classification.

Authors: Suju Rajan, Joydeep Ghosh, Melba M. Crawford.

Description: Obtaining training data for land cover classification using remotely sensed data is time consuming and expensive especially for relatively inaccessible locations. Therefore, designing classifiers that use as few labeled data points as possible is highly desirable. Existing approaches typically make use of small-sample techniques and semisupervision to deal with the lack of labeled data. In this paper, we propose an active learning technique that efficiently updates existing classifiers by using fewer labeled data points than semisupervised methods. Further, unlike semisupervised methods, our proposed technique is well suited for learning or adapting classifiers when there is substantial change in the spectral signatures between labeled and unlabeled data. Thus, our active learning approach is also useful for classifying a series of spatially/temporally related images, wherein the

spectral signatures vary across the images. Our interleaved semisupervised active learning method was tested on both single and spatially/temporally related hyperspectral data sets. We present empirical results that establish the superior performance of our proposed approach versus other active learning and semisupervised methods.

2.4. SURVEY 4

Publication: IEEE Transactions on Neural Networks - A fast and accurate online sequential learning algorithm for feed forward networks.

Authors: N.Y. Liang, Guang-bin Huang, P. Saratchandran, N. Sundararajan.

Description: In this paper, we develop an online sequential learning algorithm for single hidden layer feedforward networks (SLFNs) with additive or radial basis function (RBF) hidden nodes in a unified framework. The algorithm is referred to as online sequential extreme learning machine (OS-ELM) and can learn data one-by-one or chunk-by-chunk (a block of data) with fixed or varying chunk size. The activation functions for additive nodes in OS-ELM can be any bounded non constant piecewise continuous functions and the activation functions for RBF nodes can be any integrable piecewise continuous functions. In OS-ELM, the parameters of hidden nodes (the input weights and biases of additive nodes or the centers and impact factors of RBF nodes) are randomly selected and the output weights are analytically determined based on the sequentially arriving data. The algorithm uses the ideas of ELM of Huang developed for batch learning which has been shown to be extremely fast with generalization performance better than other batch training methods. Apart from selecting the number of hidden nodes, no other control parameters have to be manually chosen. Detailed performance comparison of OS-ELM is done with other popular sequential learning algorithms on benchmark problems drawn from the regression, classification and time series prediction areas. The results show that the OS-ELM is faster than the other sequential algorithms and produces better generalization performance.

2.5. SURVEY 5

Publication: IEEE Transactions on Speech and Audio Processing-Active learning: Theory and applications to automatic speech recognition.

Authors: G. Riccardi, D. Hakkani-Tur.

Description: We are interested in the problem of adaptive learning in the context of automatic speech recognition (ASR). In this paper, we propose an active learning algorithm for ASR. Automatic speech recognition systems are trained using human supervision to provide transcriptions of speech utterances. The goal of Active Learning is to minimize the human supervision for training acoustic and language models and to maximize the performance given the transcribed and untranscribed data. Active learning aims at reducing the number of training examples to be labeled by automatically processing the unlabeled examples, and then selecting the most informative ones with respect to a given cost function for a human to label. In this paper we describe how to estimate the confidence score for each utterance through an on-line algorithm using the lattice output of a speech recognizer. The utterance scores are filtered through the informativeness function and an optimal subset of training samples is selected. The active learning algorithm has been applied to both batch and on-line learning scheme and we have experimented with different selective sampling algorithms. Our experiments show that by using active learning the amount of labeled data needed for a given word accuracy can be reduced by more than 60% with respect to random sampling.

2.6. SURVEY 6

Publication: IEEE Trans. Neural Networks - Statistical active learning in multilayer perceptrons.

Author: K. Fukumizu.

Description: We are interested in the problem of adaptive learning in the context of automatic speech recognition (ASR). In this paper, we propose an active learning algorithm for ASR. Automatic speech recognition systems are trained using human supervision to provide transcriptions of speech utterances. The goal of Active

Learning is to minimize the human supervision for training acoustic and language models and to maximize the performance given the transcribed and untranscribed data. Active learning aims at reducing the number of training examples to be labeled by automatically processing the unlabeled examples, and then selecting the most informative ones with respect to a given cost function for a human to label. In this paper we describe how to estimate the confidence score for each utterance through an on-line algorithm using the lattice output of a speech recognizer. The utterance scores are filtered through the informativeness function and an optimal subset of training samples is selected. The active learning algorithm has been applied to both batch and on-line learning scheme and we have experimented with different selective sampling algorithms. Our experiments show that by using active learning the amount of labeled data needed for a given word accuracy can be reduced by more than 60% with respect to random sampling.

CHAPTER 3

METHODOLOGY

3.1. SYSTEM ANALYSIS

After analyzing the requirements of the task to be performed, the next step is to analyze the problem and understand its context. The first activity in the phase is studying the existing system and another is to understand the requirements and domain of the new system. Both the activities are equally important, but the first activity serves as a basis of giving the functional specifications and then the successful design of the proposed system. Understanding the properties and requirements of a new system is more difficult and requires creative thinking and understanding of existing running system is also difficult, improper understanding of the present system can lead diversion from solution.

3.1.1. Analysis Model

The model that is basically being followed is the SPIRAL MODEL, which states that the phases are organized in a linear order. First of all the feasibility study is done. Once that part is over the requirement analysis and project planning begins. If system exists one and modification and addition of the new module is needed, analysis of the present system can be used as the basic model.

The design starts after the requirement analysis is complete and the coding begins after the design is complete. Once the programming is completed, the testing is done. In this model the sequence of activities performed in a software development project is: -

- Requirement Analysis
- Project Planning
- System design
- Detail design
- Coding

- Unit System integration & testing
- Here the linear ordering of these activities is critical. End of the phase and the output of testing. One phase is the input of another phase. The output of each phase is to be consistent with the overall requirement of the system.

SPIRAL MODEL was defined by Barry Boehm in his 1988 article, “A spiral Model of Software Development and Enhancement. This model was not the first model to discuss iterative development, but it was the first model to explain why the iteration models.

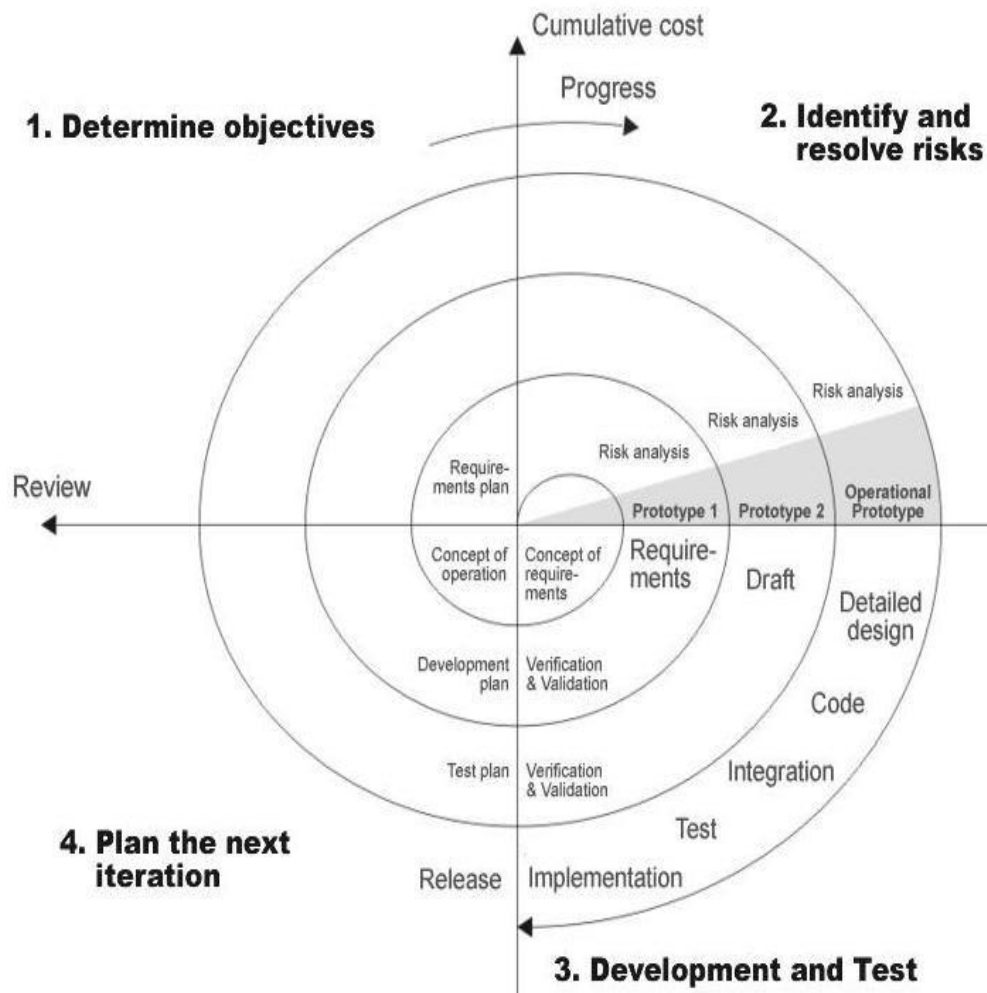


Fig 3.1 Spiral Model

3.2. EXISTING SYSTEM

Imbalanced data typically refers to classification tasks where the classes are not represented equally. Most machine learning algorithms work best when the number of samples in each class is about equal in representation of some particular data. This is because most algorithms are designed to maximize accuracy and reduce error in this application with the use of optimization in novel online sequential weighted extreme learning machine algorithms with iterative procedure that extracts more instances belonging to minor data's in this application. A batch of data that has been uploaded without the label then by splitting up of data on basis of that weightage of data is separated and labeled according to the type of data .when a set of data is generated in cluster technique by getting those data the clustered form is deployed by the new label of data. In the present system the problem of active learning in class imbalance scenario of online weighted extreme learning machine, they find that harmfulness of skewed data distribution is related to multiple factors Datasets are stored in two sets of data with labeled form and un-labeled data which has stored in the database. It has low imbalanced ratio and small classification overlapping, we have to assign a large threshold for them to construct classification models, and readers are suggested to examine the instance distribution on the initially labeled set.

3.2.1. *Disadvantage*

- It shouldn't be messy, because you don't want to spend a lot of time in cleaning and overlapping of data in the dataset.
- Accurate data are not evaluated in this process when unlabeled data are extracted from the uploaded function.
- Data sets are not appropriate with the type of data that enhances the disjunction of collected data.
- Multiple Dataset values are not labeled and identified with the set of datasets from different numbers of instances.

- It increases the likelihood of over fitting since it replicates the minority class events with reusing the dataset from the collection of particular labeled data and unlabeled data in the form of a set.

3.3. PROPOSED SYSTEM

In our proposed system learning materials are posted by public users any public persons can give their data. Data will be uploaded by a file using file upload Learning materials will be viewed by Human Annotator and he need to separate labeled and unlabelled data"s and need to allocate a label or need to continue with matching label, once the data are labeled it will be available for learners. This process is achieved by AOW-ELM algorithm. Data set is a collection of subsets of the data such that every data point is in at least one of the subsets. Formally, we say a cover U_i of a data set X is satisfies the condition that for any $x \in X$, there exists at least on $U_i \in U_i$ such that $x \in U_i$. In practice, we often have that each point is contained in multiple datasets. The functionalities that has been done here with few algorithm of handling those data with Active Learning with Extreme Learning and Online extreme machine learning by using these algorithm getting the data from unlabeled set of data and labeled set of data are managed with the type of data that has been uploaded in the form of document or text with the data is further processed. By exploring the skewed data from the probabilistic distribution of those data with accurate data type. The datasets are simplified complex created from multiple data sets by labeling each data"s.

3.3.1. *Advantage*

- Theft of data or leakage of data is not found here and there is no loss of information from the given data.
- It's simple to implement and interpret with the user by getting a labeled form of data to identify its type of instance.

- It can help improve the runtime of the model and solve the memory problems in the section of adding data in a set by reducing the number of training data samples when the training data set is enormous.

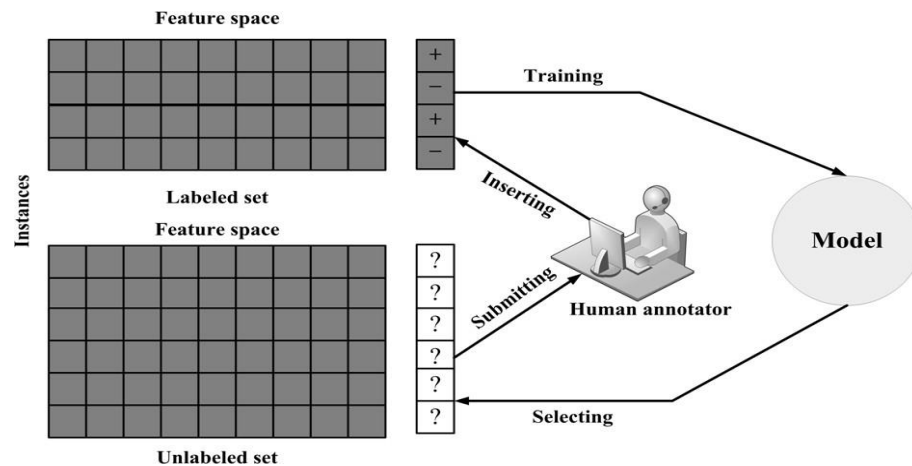


Fig 3.2 System Architecture

3.4. MODULES OF THE PROJECT

3.4.1. Authentication and Authorization

Authentication is about verification of your credentials such as Username/User ID and password to verify your personality. The system analysis whether you are using your credentials or not. Usually, authentication is done with a username and password, although there are various ways to be authenticated. However, the authorization process of this is to give access in the form of approval to the user. So that after the approval the one can able to access by verifying your rights.

3.4.2. Material Upload

The learning material will be upload by Public post, material can be uploaded by labeled or unlabeled material and all the materials will be stored in the

server, Human Annotator will collect all the labeled and unlabeled learning materials. Labeled materials will be visible in Active learning and unlabeled data will be labeled by human annotator and that material also available for Active learning.

3.4.3. Active Learning

The merits of active learning lie in decreasing both the burden of human experts and the complexity of training instances but acquiring a classification model that delivers superior or comparable performance to the model with labeling all instances. It is well known that active learning can simultaneously improve the quality of the classification model and decrease the complexity of training instances. However, several previous studies have indicated that the performance of active learning is easily disrupted by an imbalanced data distribution. Some existing imbalanced active learning approaches also suffer from either low performance or high time consumption.

3.4.4. Active Learning with Extreme Learning Machine

We proposed an active learning algorithm based on ELM and named it AOW-ELM. We found that the actual outputs of ELM can reflect the uncertainty level of instances, i.e., their classification confidences. Specifically, we also proved that there is an approximate mapping relationship between the actual outputs of ELM and the posterior probabilities in the Bayes classifier.

3.4.5. Learning Material

The material will be available for the user when all the data's are have been labeled and user can choose the learning materials from the available materials and can view the detailed learning material.

3.5. FEASIBILITY REPORT

The preliminary investigation examines project feasibility, the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging oldest running system. All system is feasible if they are unlimited

resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:

- Technical Feasibility
- Operational Feasibility
- Economic Feasibility

3.5.1. *Technical Feasibility*

The technical issue usually raised during the feasibility stage of the investigation includes the following:

- Does the necessary technology exist to do what is suggested?
- Do the proposed equipment"s have the technical capacity to hold the data required to use the new system?
- Will the proposed system provide adequate response to inquiries, regardless of the number or location of users?
- Can the system be upgraded if developed?
- Are there technical guarantees of accuracy, reliability, ease of access and data security?

Earlier no system existed to cater to the needs of „Secure Infrastructure Implementation System“. The current system developed is technically feasible. It is a browser based user interface for audit workflow. Thus, it provides an easy access to the users.

The purpose is to create, establish and maintain a workflow among various entities in order to facilitate all concerned users in their various capacities or roles. Permission to the users would be granted based on the rules specified. Therefore, it provides the technical guarantee of accuracy, reliability and security. The software and hard requirements for the development of this project are not many and are already available or are available as free as open source.

The work for the project is done with the current equipment and existing software technology. Necessary bandwidth exists for providing a fast feedback to the users irrespective of the number of users using the system.

3.5.2. *Operational Feasibility*

The analyst considers the extent the proposed system will fulfill his departments. That is, whether the proposed system covers all aspects of the working system and whether it has considerable improvements. We have found that the proposed “Secure transaction” will certainly have considerable improvements over the existing system.

3.5.3. *Economic Feasibility*

The proposed system is economically feasible because the cost involved in purchasing the hardware and the software is within approachable. Working in this system need not require a highly qualified professional. The operating-environment costs are marginal. The less time involved also helped in its economic feasibility.

3.6. SOFTWARE REQUIREMENT SPECIFICATION

The purpose of this document is to present a detailed description of the Web application system. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be proposed to the Regional Historical Society for its approval. The purpose of this Software Requirement Specification (SRS) is to help the project. It is provided with some requirements which are used in the Transaction Mercator System. All parts design, coding and testing will be prepared with helping of SRS. The purpose of this document is to detail the requirements placed on the Transaction Mercator System and serves as a contract between the customer and the developers as to what is to be expected of the stock exchange, and how the components of the system are working with each other with external systems.

3.6.1. Developers Responsibilities Overview

The developer is responsible for developing the system, which meets the SRS and solving all the requirements of the system?

- Demonstrating the system and installing the system at client's location after the acceptance testing is successful. Submitting the required user manual describing the system interfaces to work on it and also the documents of the system
- Conducting any user training that might be needed for using the system.
- Maintaining the system for a period of one year after installation.

3.6.2. Functional Requirements

Following is a list of functionalities of the browsing enabled system.

- An Activity with a UI that allows you to browser settings. Provide a second Activity that allows users to access the share with permission from the administrator. Handle activity lifecycle appropriately. A precondition for any points in this part of the grade is code that compiles and runs.
- Your application should allow a user to browse the shares, buy and sell the shares with specific metadata. The assignment requires you to create a UI for browsing and a UI for integrating the two.

3.6.3. Non-Functional Requirements

- Each member should have a separate system. The system should ask the username and password to open the application. It doesn't permit to unregistered user to access the System. The system should have Role based System functions access. Approval Process has to be defined. The system should have Modular customization components so that they can be reused across the implementation.
- Secure access of confidential data (employee's details). 24 X 7 availability.
- Better component design to get better performance at peak time.
- Flexible service based architecture will be highly desirable for future extension.

3.6.4. Performance Requirements

Performance is measured in terms of the output provided by the application. Requirement specification plays an important part in the analysis of a system. Only when the required specifications are properly given, it is possible to design a system, which will fit into required environment. It rests largely in the part of the users of the existing system to give the required specifications because they are the people who finally use the system. This is because the requirements have to be known during the initial stages so that the system can be designed according to those requirements. It is very difficult to change the system once it has been designed and on the other hand designing a system, which does not cater to the requirements of the user, is of no use.

3.7. THE .NET FRAMEWORK ARCHITECTURE

.NET Framework (pronounced dot net) is a software framework developed by Microsoft that runs primarily on Microsoft Windows. It includes a large class library known as Framework Class Library (FCL) and provides language interoperability (each language can use code written in other languages) across several programming languages. Programs written for the .NET Framework execute in a software environment (as contrasted to hardware environment), known as the Common Language Runtime (CLR), an application virtual machine that provides services such as security, memory management, and exception handling. FCL and the CLR together constitute .NET Framework.

FCL provides user interface, data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. Programmers produce software by combining their own source code with the .NET Framework and other libraries. .NET Framework is intended to be used by most new applications created for the Windows platform. Microsoft also produces an integrated development environment largely for .NET software called Visual Studio.

The first level of the representation is the operating system; the .NET layer is located between the system and applications. The second level is the

Common Language Runtime (CLR), which provides the part of the .NET Framework doing the most work. We discuss the CLR later in this chapter. The next level is the Base Class Library (BCL), which provides all .NET objects that can be used both in your code and by Visual Basic when creating applications. The BCL also provides the infrastructure of several .NET technologies that you use in building applications, such as WPF, Windows Forms, ASP.NET, WCF, and so on. The last level is represented by applications that rely on the previous layers.

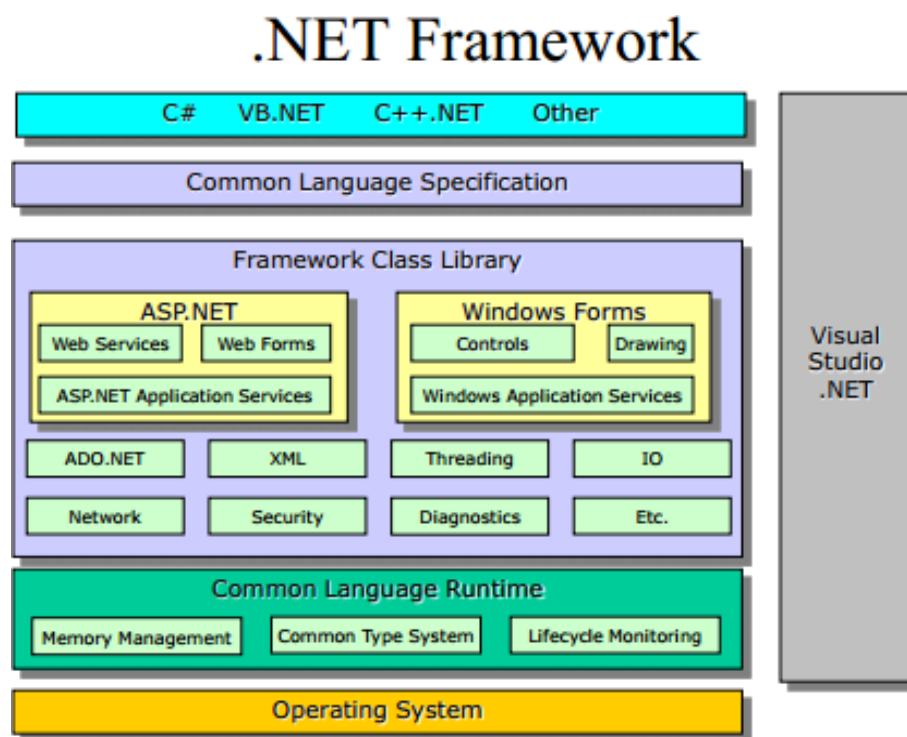


Fig 3.3 .NET Framework

3.7.1. Common Language Runtime engine

Common Language Runtime (CLR) serves as the execution engine of the .NET Framework. All .NET programs execute under the supervision of the CLR, guaranteeing certain properties and behaviors in the areas of memory management, security, and exception handling.

3.7.2. *Language Independence*

.NET Framework introduces a Common Type System, or CTS. The CTS specification defines all possible data types and programming constructs supported by the CLR and how they may or may not interact with each other conforming to the Common Language Infrastructure (CLI) specification. Because of this feature, the .NET Framework supports the exchange of types and object instances between libraries and applications written using any conforming .NET language.

3.7.3. *Framework Class Library*

Framework Class Library (FCL) is a library of functionality available to all languages using the .NET Framework. FCL provides classes that encapsulate a number of common functions, including file reading and writing, graphic rendering, database interaction, XML document manipulation, and so on. It consists of classes, interfaces of reusable types that integrate CLR.

3.7.4. *Simplified deployment*

The .NET Framework includes design features and tools which help manage the installation of computer software to ensure that it does not interfere with previously installed software, and that it conforms to security requirements.

3.7.5. *Security*

The design addresses some of the vulnerabilities, such as buffer overflows, which have been exploited by malicious software. Additionally, .NET provides a common security model for all applications.

3.7.6. *Portability*

While Microsoft has never implemented the full framework on any system except Microsoft Windows, it has engineered the framework to be platform-agnostic, and cross-platform implementations are available for other operating systems (see the Silver light and Alternative implementations). Microsoft submitted the specifications for CLI (which includes the core class libraries, CTS, and the Common Intermediate Language), C#, [20] and C++/CLI to both ECMA and ISO, making them available as official standards. This makes it possible for third parties to

create compatible implementations of the framework and its languages on other platforms.

3.7.7. Common Language Specification

CLS is a set of rules that specifies the features that all languages should support. CLS is an agreement among language designers and class library designers about the features and usage conventions that can be relied upon. Since some languages are not case sensitive. This does not mean all languages are not case sensitive above the CLR.

Goal: have the .NET framework support multiple languages.

Example: public names should not rely on case for uniqueness

3.8. SQL SERVER 2014

SQL is a relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases. It is a popular choice of database for the use in web applications and is an open source product. The process of setting up a SQL database varies from host to host, however, we will end up with a database name, a user name and a password. Before using our database, we must create a table. A table is a section of the database for storing related information. In a table, we will set up the different fields that will be used. Creating a table is simple; we just have to type the name, select the number of fields and click the „go“ button. We will then be taken to a setup screen where you must create the fields in the database.

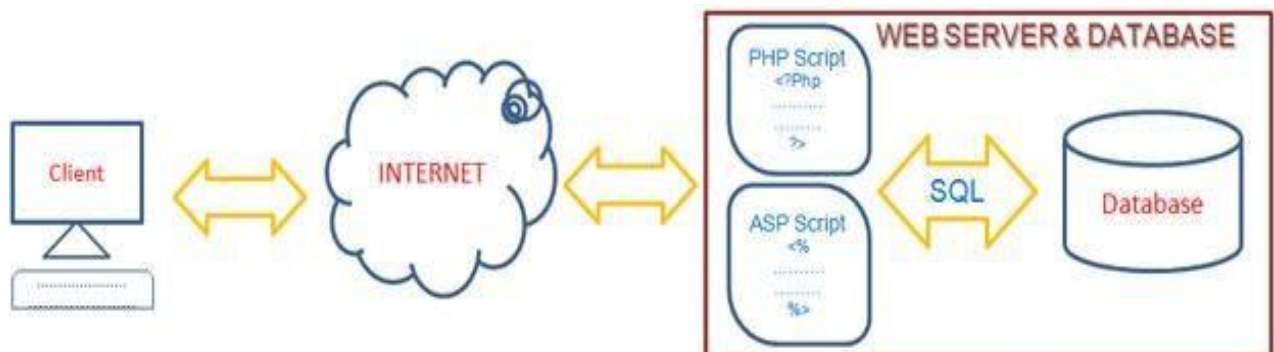


Fig 3.4 Server Architecture

SQL is used to communicate with a database. According to ANSI, it is the standard language for relational database management systems. SQL statements are used to perform tasks such as update data on a database, or retrieve data from a database. Some common relational database management systems that use SQL are: Oracle, Sybase, Microsoft SQL Server, Access, Ingres, etc. Although most database systems use SQL, most of them also have their own additional proprietary extensions that are usually only used on their system. However, the standard SQL commands such as "Select", "Insert", "Update", "Delete", "Create", and "Drop" can be used to accomplish almost everything that one needs to do with a database.

Microsoft SQL Server is a relational database management system developed by Microsoft. As a database, it is a software product whose primary function is to store and retrieve data as requested by other software applications, be it those on the same computer or those running on another computer across a network (including the Internet). SQL Server 2014 (formerly codenamed "Katmai") was released on April 2014, and aims to make data management, self-tuning, self organizing, and self maintaining with the development of SQL Server Always On technologies, to provide near-zero downtime. SQL Server 2014 also includes support for structured and semi-structured data, including digital media formats for pictures, audio, video and other multimedia data. In current versions, such multimedia data can be stored as BLOBs, but they are generic bit streams. Intrinsic awareness of multimedia data will allow specialized functions to be performed on them. According to Paul Flessner, senior Vice President, Server, Applications, Microsoft Corp., SQL Server 2014 can be a data storage backend for different varieties of data: XML, email, time/calendar, file, document, spatial, etc as well as perform search, query, analysis, sharing, and synchronization across all data types.

3.9. SYSTEM DESIGN

Software design sits in the technical kernel of the software engineering process and is applied regardless of the development paradigm and area of application. Design is the first step in the development phase for any engineered

product or system. The designer's goal is to produce a model or representation of an entity that will later be built. Beginning, once system requirement has been specified and analyzed, system design is the first of the three technical activities -design, code and test that is required to build and verify software. The importance can be stated with a single word "Quality". Design is the place where quality is fostered in software development. Design provides us with representations of software that can assess for quality. Design is the only way that we can accurately translate a customer's view into a finished software product or system. Software design serves as a foundation for all the software engineering steps that follow. Without a strong design we risk building an unstable system – one that will be difficult to test, one whose quality cannot be assessed until the last stage. During design, progressive refinement of data structure, program structure, and procedural details are developed reviewed and documented. System design can be viewed from either technical or project management perspective. From the technical point of view, design is comprised of four activities – architectural design, data structure design, interface design and procedural design.

3.10. SYSTEM TESTING AND IMPLEMENTATION

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and coding. In fact, testing is the one step in the software engineering process that could be viewed as destructive rather than constructive.

A strategy for software testing, integrates software test case design methods into a well-planned series of steps that result in the successful construction of software. Testing is the set of activities that can be planned in advance and conducted systematically. The underlying motivation of program testing is to affirm software quality with methods that can economically and effectively apply both strategic to both large and small-scale systems.

3.10.1. *Strategic Approach to Software Testing*

The software engineering process can be viewed as a spiral. Initially system engineering defines the role of software and leads to software requirement

analysis where the information domain, functions, behavior, performance, constraints and validation criteria for software are established. Moving inward along the spiral, we come to design and finally to coding. To develop computer software we spiral in along streamlines that decrease the level of abstraction at each turn.

A strategy for software testing may also be viewed in the context of the spiral. Unit testing begins at the vertex of the spiral and concentrates on each unit of the software as implemented in source code. Testing progress is done by moving outward along the spiral to integration testing, where the focus is on the design and the construction of the software architecture. Talking another turn on outward on the spiral we encounter validation testing where requirements established as part of software requirements analysis are validated against the software that has been constructed. Finally, we arrive at system testing, where the software and other system elements are tested as a whole.

3.11. SYSTEM SECURITY

The protection of computer based resources that include hardware, software, data, procedures and people against unauthorized use or natural Disaster is known as System Security.

Security system can be divided into four related issues:

- Security
- Integrity
- Privacy
- Confidentiality

3.11.1. *Security in Software*

System security refers to various validations on data in the form of checks and controls to avoid the system from failing. It is always important to ensure that only valid data is entered and only valid operations are performed on the system. The system employs two types of checks and controls:

3.11.1.1. Client Side Validation

Various client side validations are used to ensure on the client side that only valid data is entered. Client side validation saves server time and load to handle invalid data. Some checks are imposed:

- JavaScript is used to ensure those required fields are filled with suitable data only. Maximum lengths of the fields of the forms are appropriately defined.
- Forms cannot be submitted without filling up the mandatory data so that manual mistakes of submitting empty fields that are mandatory can be sorted out at the client side to save the server time and load.
- Tab-indexes are set according to the need and taking into account the ease of use while working with the system.

3.11.1.2. Server Side Validation

Some checks cannot be applied on the client side. Server side checks are necessary to save the system from failing and intimating the user that some invalid operation has been performed or the performed operation is restricted. Some of the server side checks imposed is:

- A server side constraint has been imposed to check for the validity of primary key and foreign key. A primary key value cannot be duplicated. Any attempt to duplicate the primary value results in a message intimating the user about those values through the forms using foreign key can be updated only of the existing foreign key values.
- The user is intimated through appropriate messages about the successful operations or exceptions occurring at server side.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1. NORMALIZATION

It is a process of converting a relation to a standard form. The process is used to handle the problems that can arise due to data redundancy, i.e. repetition of data in the database, maintain data integrity as well as handling problems that can arise due to insert, update, delete anomalies. Decomposing is the process of splitting relations into multiple relations to eliminate anomalies and maintain anomalies and maintain data integrity. To do this we use normal forms or rules for structuring relation.

Table 4.1 Design for User Register

S.no	Logical name	Data type	Is Null	Comments
1	ID	INT	N	This is primary column, used to identify the row uniquely.
2	Username	VARCHAR(20)	N	Denotes the original name of the user.
3	Email ID	VARCHAR(20)	N	Used to confirm the identity of a user.

4	Mobile No	VARCHAR(25)	N	Used to communicate with the user.
5	Age	VARCHAR(MAX)	N	Denotes Age of user.
6	Security Question	VARCHAR(50)	N	Authentication Process of User
7	Security Answer	VARCHAR(50)	N	Used to confirm the Authorisation of user.
8	Date	DATE	N	Get the current Date
9	Security Key	INT	N	Used to confirm the identity of a user
10	Block Status	VARCHAR(50)	N	Used to find Account status
11	Password	VARCHAR(50)	N	Used to confirm the identity of a user

Table 4.2 File Upload

S.no	Logical name	Data type	Is Null	Comments
1	File Id	INT	N	This is primary column, used to identify the row uniquely.
2	Subject	VARCHAR(50)	N	Used to get the subject of Files
3	File Path	VARCHAR(MAX)	N	Used to get the Path of Files
4	Bytes	BYTES	N	Used to convert files to bytes
5	File Data	VARCHAR(MAX)	N	Used to get the Content of the files
6	File Name	VARCHAR(50)	N	Used to get File Name
7	File Extension	VARCHAR(50)	N	Used to identify the file Extension

8	File Size	VARCHAR(50)	N	Used to get size of the File in KB
9	File Verify	VARCHAR(50)	N	Used to verify the uploaded files
10	Labelled By	VARCHAR(50)	N	Used to label the uploaded files

4.2. E – R DIAGRAMS

The relation upon the system is structured through a conceptual ER-Diagram, which not only specifies the existing entities, but also the standard relations through which the system exists and the cardinalities that are necessary for the system state to continue. The Entity Relationship Diagram (ERD) depicts the relationship between the data objects. The ERD is the notation that is used to conduct, the data modeling activity the attributes of each data object noted, is the ERD can be described design a data object description.

The set of primary components that are identified by the ERD are

- Data object
- Relationships
- Attributes
- Various types of indicators.

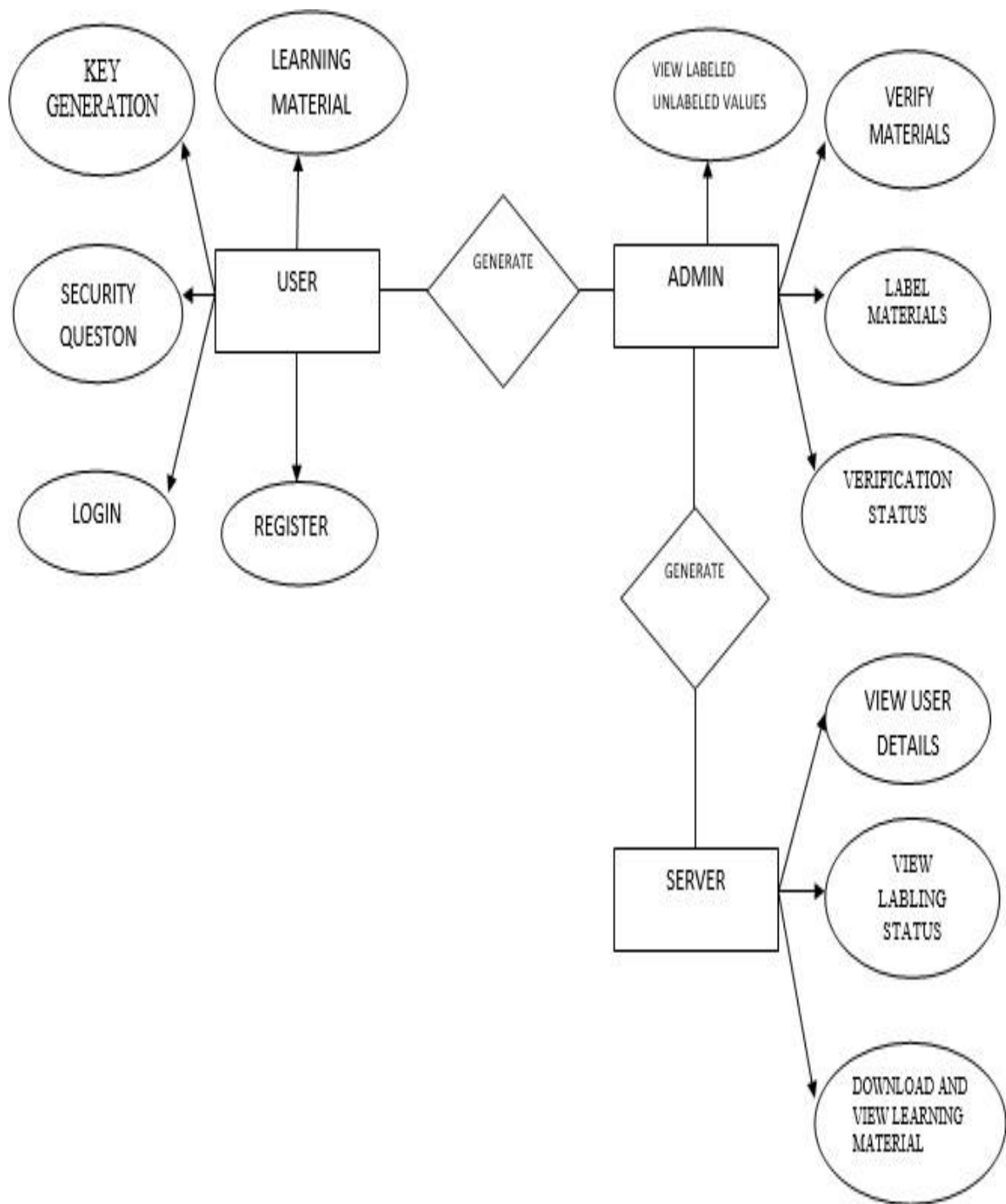


Fig 4.1 ER Diagram for Human Annotator

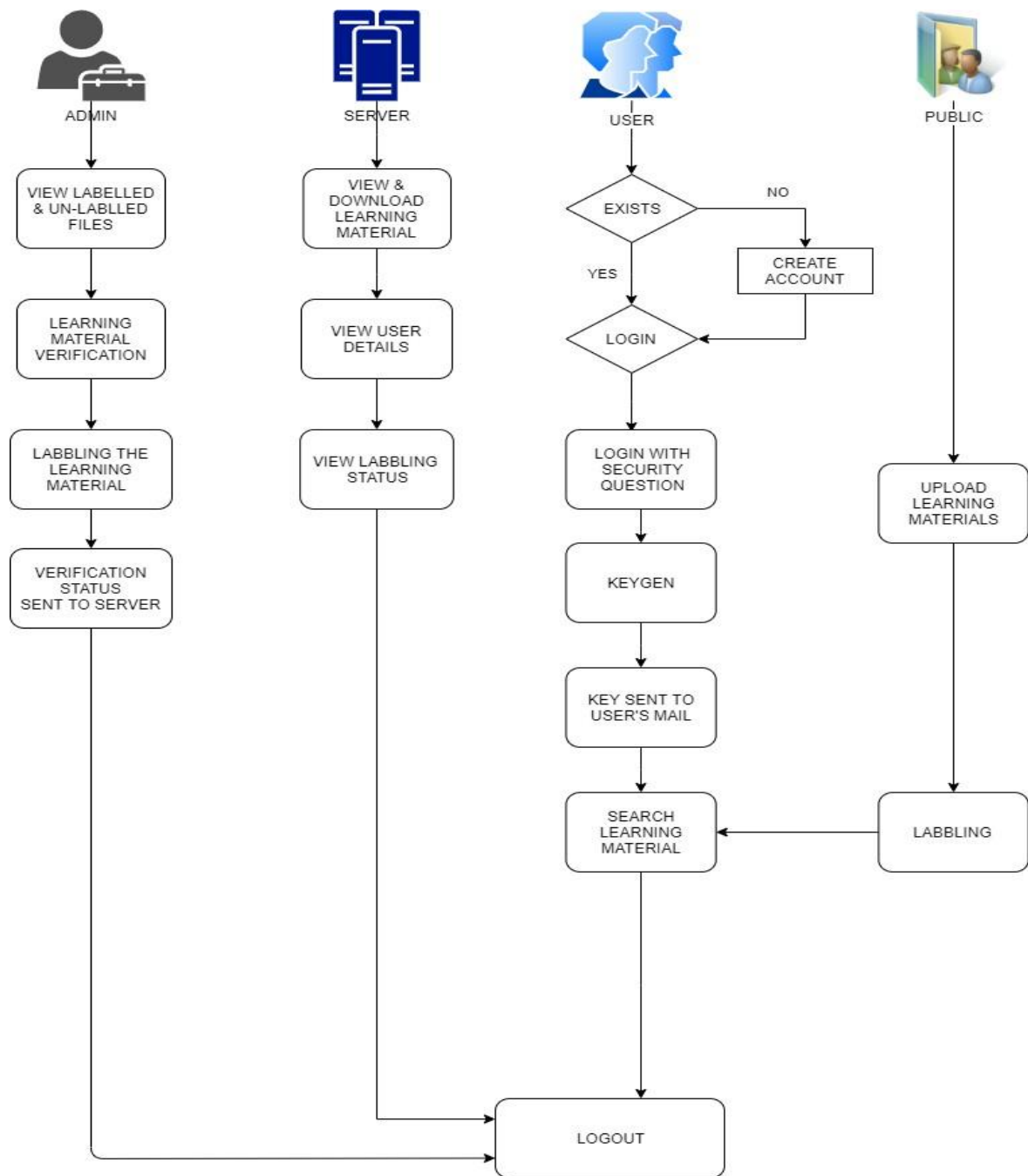


Fig 4.2 Flow Chart of Human Annotator

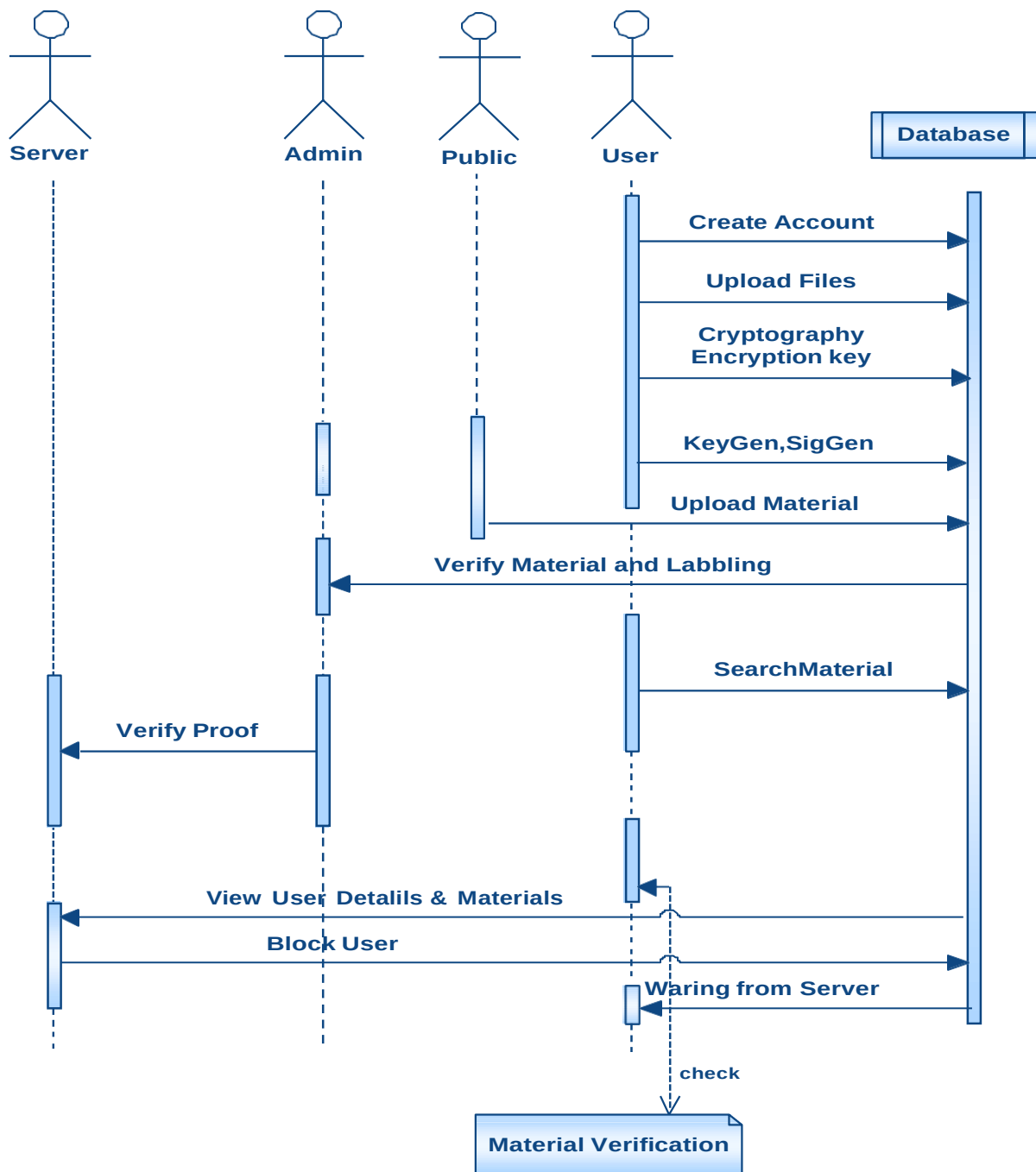


Fig 4.3 Sequence Diagram for Human Annotator

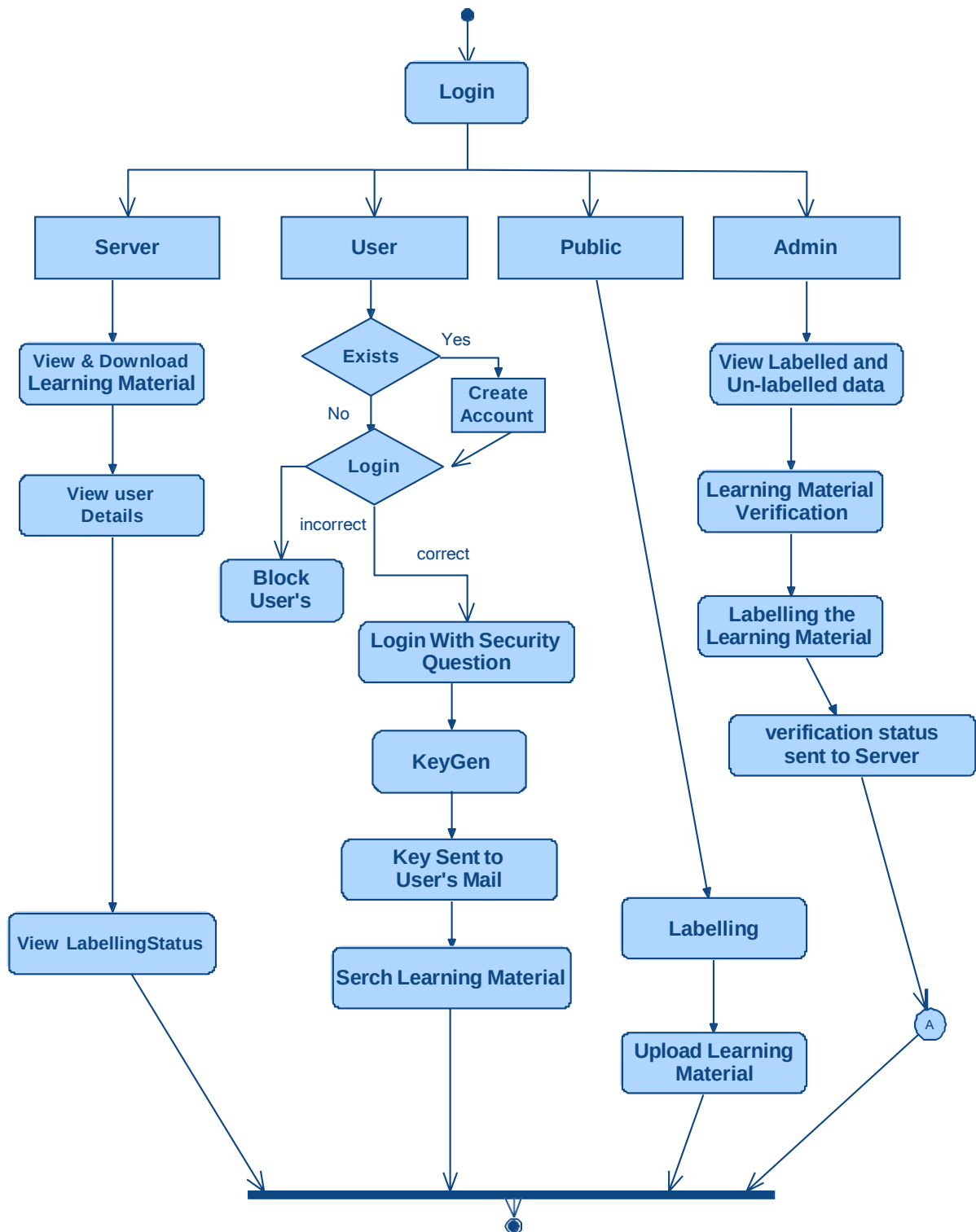


Fig 4.4 Activity Diagram for Human Annotator

4.3. DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a graphical tool used to describe and analyze movement of data through system. It is a graphical representation of the “flow” of data through a computer system or a data or it looks at how data flows through a system. These are central tool and basic from which the other components are developed. The transformation of data from input to output, through processed may be described logically and independently of physical components associated with the system. The development of DFD is done at several levels. The flow diagram describes the boxes that describe computations, decisions, interactions & loops. It is important to keep in mind that the flow diagrams are not flowcharts and should not include control elements.

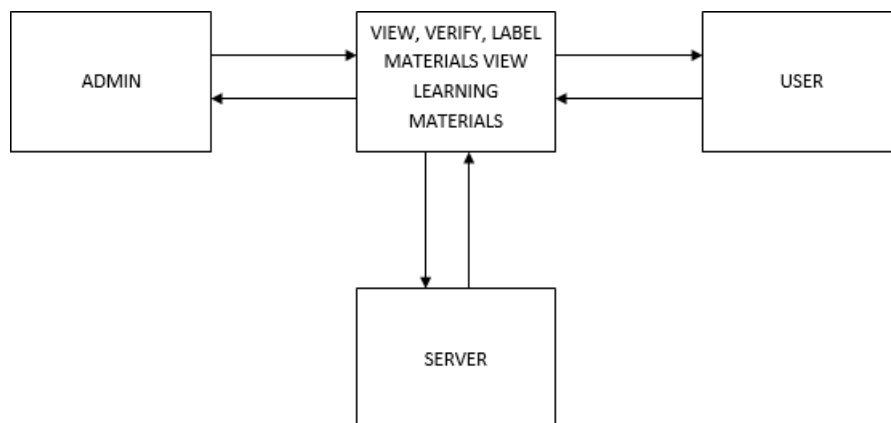


Fig 4.5 Level - 0 Data Flow Diagram

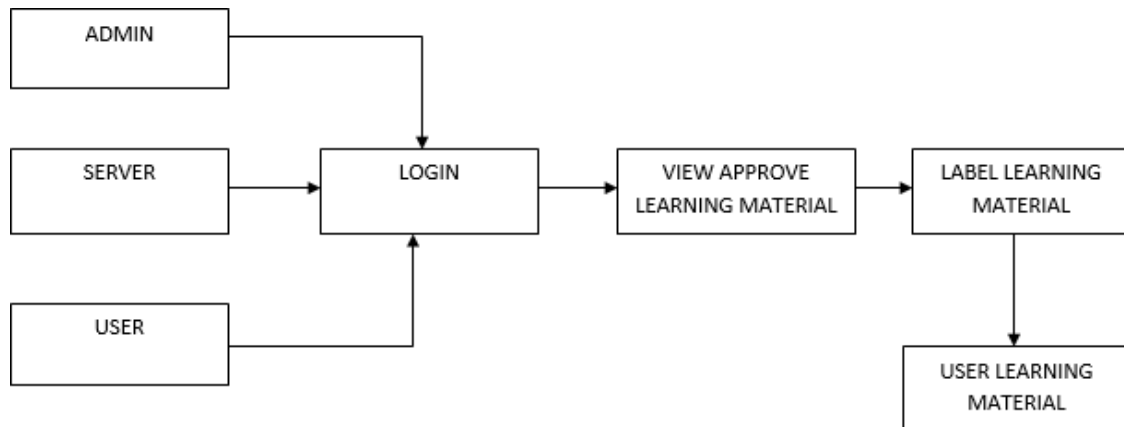


Fig 4.6 Level – 1 Data Flow Diagram

4.4. ALGORITHM

Active Online Weighted Extreme Learning Machine (AOW - ELM)

- Online Sequential Weighted Extreme Learning Machine
- Weight Update Rule
- Early Stopping Criterion
- AOW-ELM Algorithm Description

Input: A collected unlabeled instance set, the number of initially labeled instances κ , the initially labeled set

$\psi = \varphi$, the number of labeled instances on each round

u , the number of the hidden nodes L , the penalty factor

C , and the stopping threshold μ .

Output: The final weight matrix $\beta(F)$.

Overall, 32 binary-class imbalanced data sets that are used, most of which come from the University of California-Irvine (UCI) machine learning data repository, and the others are from several publications about bioinformatics. These data sets have different number of instances, number of features, and class imbalance ratios. In addition, for each data set, half of the instances are reserved as the test set, and in the other half, a specific percentage of instances is extracted as

an initially labeled seed set. The percentage correlates with the size of the data set and class imbalance ratio to avoid cold starts as much as possible. Furthermore, the batch size of active learning, i.e., the number of instances labeled on each round, is also associated with the size of data set.

EL Machine was initially bounced for single layer feed forward systems anywhere info loads of a SLFN are haphazardly produced, and yield loads are prepared with bunch learning procedure of less squares. It has been demonstrated haphazardly shrouded neurons and tunable yield loads having all-inclusive estimation and astounding speculation execution. All the more critically, ELM outflanks most existed learning calculations in preparing speed and it has been broadly utilized in utilizations of face acknowledgment, picture handling and classification, power value arrangement, vitality product prospects record estimate, area fingerprinting strategy, protein succession order, and area grouping. Data streams with class awkwardness happen as a rule in numerous genuine applications.

Online consecutive learning is the one of the successful strategies for ordering information creek with class irregularity. Based on online successive extraordinary one, the proposed dw-ELM strategy breaks down the circulation normal for information taking into account existence, and gives a versatile double weighting plan to tune the loads at both the time level and the space level. Broad trial assessments on 10 imbalanced datasets demonstrate that the technique beats a few looking at strategies as far as G-mean and F-measure measurements. Imbalanced information arrangement by consolidating casting a ballot technique into weight as outrageous learning machine (weighted ELM), these significant advances are associated with the proposed strategy: preparing weighted ELM classifiers, killing not usable orders to decide appropriate classifiers for casting a ballot, lastly deciding the characterization result dependent on greater part casting a ballot.

Human Annotator will gather all the named and unlabeled learning materials named materials will be noticeable in Active learning and unlabeled information will be marked by human annotator and that material additionally accessible for Active learning. We proposed an AOW-ELM; In particular, we additionally demonstrated that there is an inexact mapping connection between the

genuine yields of ELM also probabilities in the Bayes classifier. The material will be accessible for the client when all the information's are have been named and client can pick the taking in materials from the accessible materials and can see the definite learning material and materials can be downloaded for their utilization.

The merits of the AOW-ELM algorithm can be summarized as follows.

- It has a robust weight update rule.
- Its running time is fast and linear with the training instances.
- It has a flexible early stopping criterion.
- It is appropriate for various types of data sets.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The labeled and unlabelled data have been analyzed by human annotators and matched with labeled appropriate data sets are achieved and it will be available for learners. The clustering of data is shared from the user by getting the instance with proper updating of data. In future enhancement sample and clustering must be separated into material and evidence at the time of posting public data"s, By Recall method, recall is the number of True Positives divided by the number of True Positives and the number of False Negatives. By the data that has been posted in the form of sudation with document in that particular data that is labeled and unlabelled. Those unlabelled data are separated by type of data and by letting the data with a set of labeled data. By adding another way it is the number of positive predictions divided by the number of positive class values in the test data. It is also called Sensitivity or the True Positive Rate. In the future work, we will focus more on the problem of active learning on multiclass imbalanced data sets. In addition, the active learning strategies addressing imbalanced and unlabeled data streams with handling concept drifts will also be investigated.

REFERENCES

1. G. Riccardi and D. Hakkani-Tur, "Active learning: Theory and applications to automatic speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, Jul. 2005, pp. 504-511.
2. H. Liao, L. Chen, Y. Song, and H. Ming, "Visualization-based active learning for video annotation," *IEEE Trans. Multimedia*, vol. 18, no. 11, Nov. 2016, pp. 2196-2205.
3. K. Fukumizu, "Statistical active learning in multilayer perceptrons," *IEEE Trans. Neural Networks.*, vol. 11, no. 1, Jan. 2000, pp. 17-26.
4. M. Wang and X. S. Hua, "Active learning in multimedia annotation and retrieval: A survey," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, February 2011, p. 10.
5. N.Y. Liang, G.B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. Neural Networks.*, vol. 17, no. 6, Nov. 2006, pp. 1411-1423.
6. P. H. Gosselin and M. Cord, "Active learning methods for interactive image retrieval," *IEEE Trans. Image Process.*, vol. 17, no. 7, Jul. 2008, pp. 1200-1211.
7. S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, April 2008, pp. 1231-1242.

APPENDIX

A. SAMPLE CODE

AdminLogin.aspx.cs

```
using System;
using System.Collections;
using System.Configuration;
using System.Data;
using System.Linq;
using System.Web;
using System.Web.Security;
using System.Web.UI;
using System.Web.UI.HtmlControls;
using System.Web.UI.WebControls;
using System.Web.UI.WebControls.WebParts;
using System.Xml.Linq;
public partial class AdminLogin : System.Web.UI.Page
{
    string upid, uppwd;
    protected void Page_Load(object sender, EventArgs e)
    {
    }
    protected void ImageButton1_Click(object sender, ImageClickEventArgs e)
    {
        Response.Redirect("Default.aspx");
    }
    protected void ImageButton1_Click1(object sender, ImageClickEventArgs e)
    {
        upid = TextBox1.Text.ToUpper();
        uppwd = TextBox2.Text.ToUpper();
    }
}
```

```

if (TextBox1.Text == "")
{
    string myStringVariable1 = string.Empty;
    myStringVariable1 = "Enter Admin ID";
    ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('" +
myStringVariable1 + "');", true);
}
else
{
    if (TextBox2.Text == "")
    {
        string myStringVariable1 = string.Empty;
        myStringVariable1 = "Enter Admin Password";
        ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('"
+ myStringVariable1 + "');", true);
    }
    else
    {
        if (upid == "SERVER" && uppwd == "SERVER")
        {
            Response.Redirect("AdminMain.aspx");
            //Session["AdminID"] = TextBox1.Text;
        }
        else
        {
            string myStringVariable1 = string.Empty;
            myStringVariable1 = "Enter Admin ID/Password Correctly.";
            ClientScript.RegisterStartupScript(this.GetType(), "myalert",
"alert('" + myStringVariable1 + "');", true);
        }
    }
}

```

```

    }
}
protected void ImageButton2_Click(object sender, ImageClickEventArgs e)
{
    TextBox1.Text = "";
    TextBox2.Text = "";
}
}

```

Labelling1.aspx.cs

```

public partial class Labbling1 : System.Web.UI.Page
{
    SqlConnection con = new
SqlConnection(ConfigurationManager.AppSettings["ConnectionString"]);
    Cryptography cr = new Cryptography();
    string fileid, fverify = "YES";
    protected void Page_Load(object sender, EventArgs e)
    {
        fileid = (string)Request.Params["ID"];
        Session["FileID"] = fileid;
        SqlDataAdapter adp = new SqlDataAdapter("Select * from filearchive
where fid='" + fileid + "'", con);
        DataSet ds = new DataSet();
        adp.Fill(ds);
        Label4.Text = ds.Tables[0].Rows[0]["fid"].ToString();
        Label7.Text = ds.Tables[0].Rows[0]["ffilename"].ToString();
        Label10.Text = ds.Tables[0].Rows[0]["fsubject"].ToString();
        Label13.Text = ds.Tables[0].Rows[0]["fext"].ToString();
        Label16.Text = ds.Tables[0].Rows[0]["fsizeinkb"].ToString();
        Label19.Text = "Uploads/" + ds.Tables[0].Rows[0]["ffilename"].ToString();
        Label26.Text = ds.Tables[0].Rows[0]["fdatetime"].ToString();
    }
}

```

```

Label23.Text = ds.Tables[0].Rows[0]["labelledby"].ToString();
if (Label23.Text != "--Select--")
{
    ImageButton1.Visible = false;
    DropDownList1.Visible = false;
    Label23.Visible = true;
    string myStringVariable1 = string.Empty;
    myStringVariable1 = "Material already Labelled.";
    ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('" +
myStringVariable1 + "');", true);
}
else
{
    DropDownList1.Visible = true;
    ImageButton1.Visible = true;
    Label23.Visible = false;
}
}

protected void ImageButton1_Click(object sender, ImageClickEventArgs e)
{
    if (DropDownList1.SelectedIndex == 0)
    {
        string myStringVariable1 = string.Empty;
        myStringVariable1 = "Please Select any Material Type !";
        ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('" +
myStringVariable1 + "');", true);
    }
    else
    {
        con.Open();
    }
}

```

```

        SqlCommand cmd1 = new SqlCommand("update filearchive set
labelledby='" + DropDownList1.SelectedItem.Text + "' where fid='" +
Label4.Text + "'", con);
        cmd1.ExecuteNonQuery();
        con.Close();
        Response.Redirect("Labbling2.aspx");
    }
}
}

```

UserLearning.aspx.cs

```

public partial class UserLearning : System.Web.UI.Page
{
    SqlConnection con = new
SqlConnection(ConfigurationManager.AppSettings["ConnectionString"]);
    Cryptography cr = new Cryptography();
    string fileid, fverify = "YES";
    protected void ImageButton1_Click(object sender, ImageClickEventArgs e)
    {
        TextBox1.Text = "";
        if (DropDownList1.SelectedIndex == 0)
        {
            string myStringVariable1 = string.Empty;
            myStringVariable1 = "Please Select any Learning Material Type !";
            ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('" +
myStringVariable1 + "');", true);
        }
        else
        {
            con.Open();

```

```

        SqlDataAdapter adp = new SqlDataAdapter("Select * from filearchive
where labelledby='" + DropDownList1.SelectedItem.Text + "' and
fverify='YES'", con);

        DataSet ds = new DataSet();
        adp.Fill(ds);
        if (ds.Tables[0].Rows.Count != 0)
        {
            for (int i = 0; i <= ds.Tables[0].Rows.Count - 1; i++)
            {
                TextBox1.Text = TextBox1.Text +
ds.Tables[0].Rows[i]["fileorgdata"].ToString() + "<\br>";
            }
            con.Close();
            string myStringVariable1 = string.Empty;
            myStringVariable1 = "SUCCESS ! Admin collect all the materials and
MATCH the Results !";
            ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('"
+ myStringVariable1 + "');", true);
        }
        else
        {
            string myStringVariable11 = "LEARNING MATERIAL NOT
AVAILABLE !";
            ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('"
+ myStringVariable11 + "');", true);
        }
    }
}
}

```

OwnerFileUpload.aspx.cs

```
public partial class OwnerFileUpload : System.Web.UI.Page
{
    SqlConnection con = new
SqlConnection(ConfigurationManager.AppSettings["ConnectionString"]);
    Class1 cs = new Class1();
    Cryptography cr = new Cryptography();
    string gMailAccount = "customerservice404@gmail.com";
    string to;
    string subject = "Upload Status";
    string message;
    string pwd1, sub;
    string mydocpath, filenamemain;
    static string ownerid, filePath, filename, filesize, fileext, date;
    string cryptographykey, fileenccryptographykey;
    static byte[] filebytes;
    static float filesiz = 0f;
    string splitfilesize, checkfilename, yes, yes1, binary1, binary1enc;
    string fverify = "NO", fmodify = "", fdownload = "Block", keyrequest = "Null",
emailid;
    protected void Page_Load(object sender, EventArgs e)
    {
        Label4.Text = Convert.ToString(cs.idgeneration1());
        ownerid = (string)Session["userid"];
    }
    protected void Button1_Click(object sender, EventArgs e)
    {
        TextBox2.Text = "";
        TextBox3.Visible = false;
    }
    protected void ImageButton2_Click(object sender, ImageClickEventArgs e)
```

```

{
    Label6.Visible = false;
}
protected void ImageButton3_Click(object sender, ImageClickEventArgs e)
{
    TextBox2.Text = "";
}
protected void ImageButton1_Click(object sender, ImageClickEventArgs e)
{
    filename = System.IO.Path.GetFileName(FileUpload1.FileName);
    if (filename == "")
    {
        string myStringVariable1 = string.Empty;
        myStringVariable1 = "select any one file";
        ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('" +
myStringVariable1 + "');", true);
    }
    else
    {
        filePath = Request.PhysicalApplicationPath + "Uploads/" +
System.IO.Path.GetFileName(FileUpload1.FileName);
        FileUpload1.SaveAs(filePath);
        filebytes = new byte[FileUpload1.PostedFile.InputStream.Length];
        FileUpload1.PostedFile.InputStream.Read(filebytes, 0,
filebytes.Length);
        System.IO.Path.GetExtension(FileUpload1.PostedFile.FileName);
        filesiz = (float)FileUpload1.PostedFile.ContentLength / 1024;
        filesize = Convert.ToString(filesiz) + " KB";
        Panel3.Visible = true;
        Panel2.Visible = false;
    }
}

```

```

    }
protected void ImageButton4_Click(object sender, ImageClickEventArgs e)
{
    date = Convert.ToString(DateTime.Now.ToShortDateString());
    Random val = new Random();
    int rno = val.Next(12345, 54321);
    cryptographykey = Convert.ToString(rno);
    fileencryptcryptographykey = cr.Encrypt(cryptographykey);
    splitfilesize = Convert.ToString(filesiz);
    for (int i = 0; i < splitfilesize.Length; i++)
    {
        char c = splitfilesize[i];
        string binary = Convert.ToString(c, 2);
        binary1 = binary1 + binary;
    }
    binary1enc = cr.Encrypt(binary1);
    if (TextBox2.Text == "")
    {
        string myStringVariable1 = string.Empty;
        myStringVariable1 = "Enter File Subject.";
        ClientScript.RegisterStartupScript(this.GetType(), "myalert", "alert('" +
myStringVariable1 + "');", true);
    }
    else
    {
        TextBox3.Text = binary1;
        ImageButton4.Visible = false;
        SqlDataAdapter adp = new SqlDataAdapter("select * from filearchive",
con);
        DataSet ds = new DataSet();
        adp.Fill(ds);
    }
}

```

```

for (int i = 0; i < ds.Tables[0].Rows.Count; i++)
{
    checkfilename = ds.Tables[0].Rows[i]["filename"].ToString();
    if (checkfilename == filename)
    {
        yes = "yes";
        goto Outer;
    }
    else
    {
        yes = "no";
    }
}

```

Web.config

```

<appSettings>
    <add key="ConnectionString" value="Data Source=DESKTOP-
HSHUH7V;Initial Catalog=ImbalanceData;Integrated Security=True"/>
</appSettings>

```

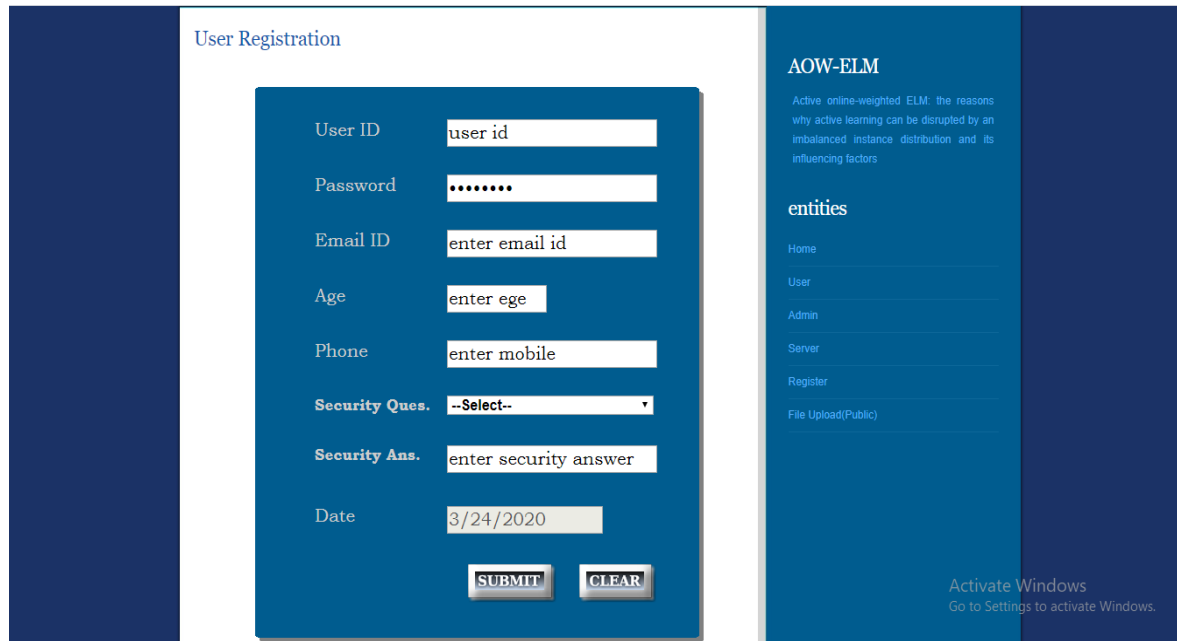
Default.aspx

```

<%@ Page Language="C#" AutoEventWireup="true"
MasterPageFile="~/MasterPage.master" CodeFile="Default.aspx.cs"
Inherits="_Default" %>
<asp:Content ContentPlaceHolderID="Default" runat="server">
    <h2 style="position: absolute; top: 210px; left: 207px;" class="title"><a
href="#">Architecture </a></h2>
    <asp:Image Style="position: absolute; top: 270px; left: 215px; width: 608px;
height: 302px;"
        ID="Image1" runat="server" ImageUrl="~/Images/arch.png" />
    &nbsp; </asp:Content>

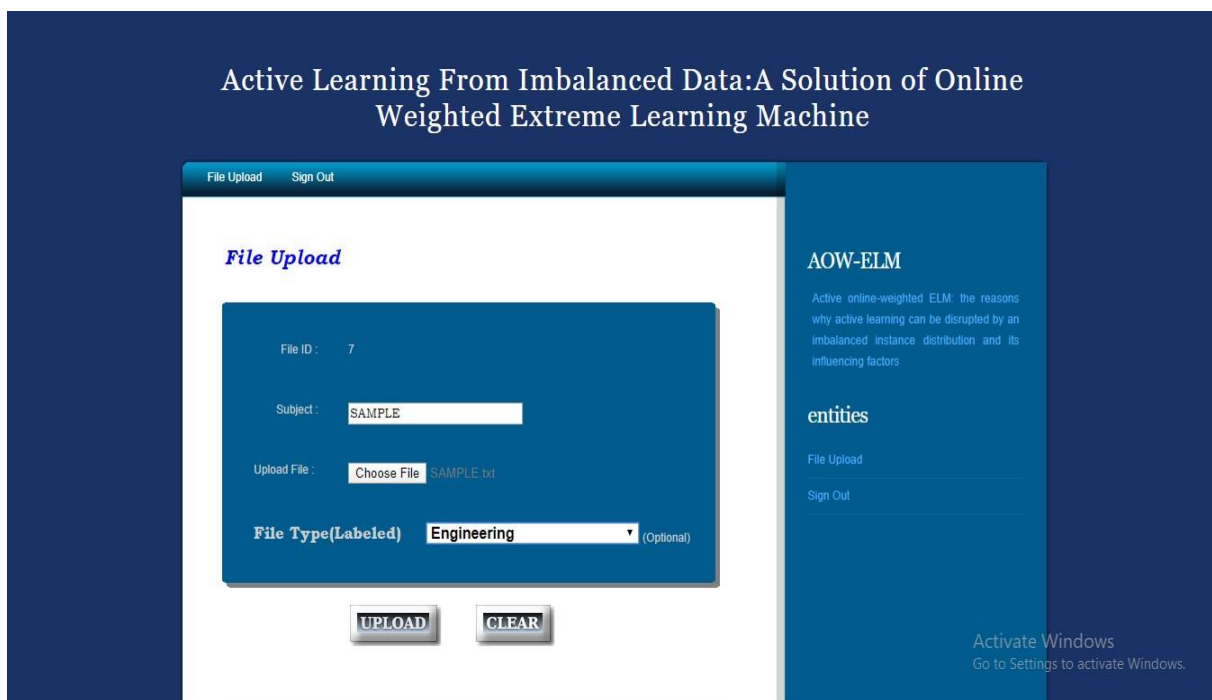
```

B. SCREENSHOTS



The screenshot displays a 'User Registration' form within a web application. The form is titled 'User Registration' and is set against a light blue background. It contains several input fields: 'User ID' (with 'user id' entered), 'Password' (masked with dots), 'Email ID' (with 'enter email id' entered), 'Age' (with 'enter ege' entered), 'Phone' (with 'enter mobile' entered), 'Security Ques.' (a dropdown menu showing '--Select--'), 'Security Ans.' (with 'enter security answer' entered), and 'Date' (with '3/24/2020' entered). At the bottom of the form are two buttons: 'SUBMIT' and 'CLEAR'. To the right of the form, there is a sidebar with the title 'AOW-ELM' and a description: 'Active online-weighted ELM: the reasons why active learning can be disrupted by an imbalanced instance distribution and its influencing factors'. Below this, there is a section titled 'entities' with a list of links: 'Home', 'User', 'Admin', 'Server', 'Register', and 'File Upload(Public)'. At the bottom right of the sidebar, there is a message: 'Activate Windows Go to Settings to activate Windows.'

Figure 1 NEW USER REGISTRATION



The screenshot displays a 'File Upload' form within a web application. The form is titled 'File Upload' and is set against a light blue background. It contains several input fields: 'File ID' (with '7' entered), 'Subject' (with 'SAMPLE' entered), 'Upload File' (with a 'Choose File' button and 'SAMPLE.txt' entered), and 'File Type(Labeled)' (a dropdown menu showing 'Engineering' and '(Optional)'). At the bottom of the form are two buttons: 'UPLOAD' and 'CLEAR'. To the right of the form, there is a sidebar with the title 'AOW-ELM' and a description: 'Active online-weighted ELM: the reasons why active learning can be disrupted by an imbalanced instance distribution and its influencing factors'. Below this, there is a section titled 'entities' with a list of links: 'File Upload' and 'Sign Out'. At the bottom right of the sidebar, there is a message: 'Activate Windows Go to Settings to activate Windows.'

Figure 2 FILE UPLOAD

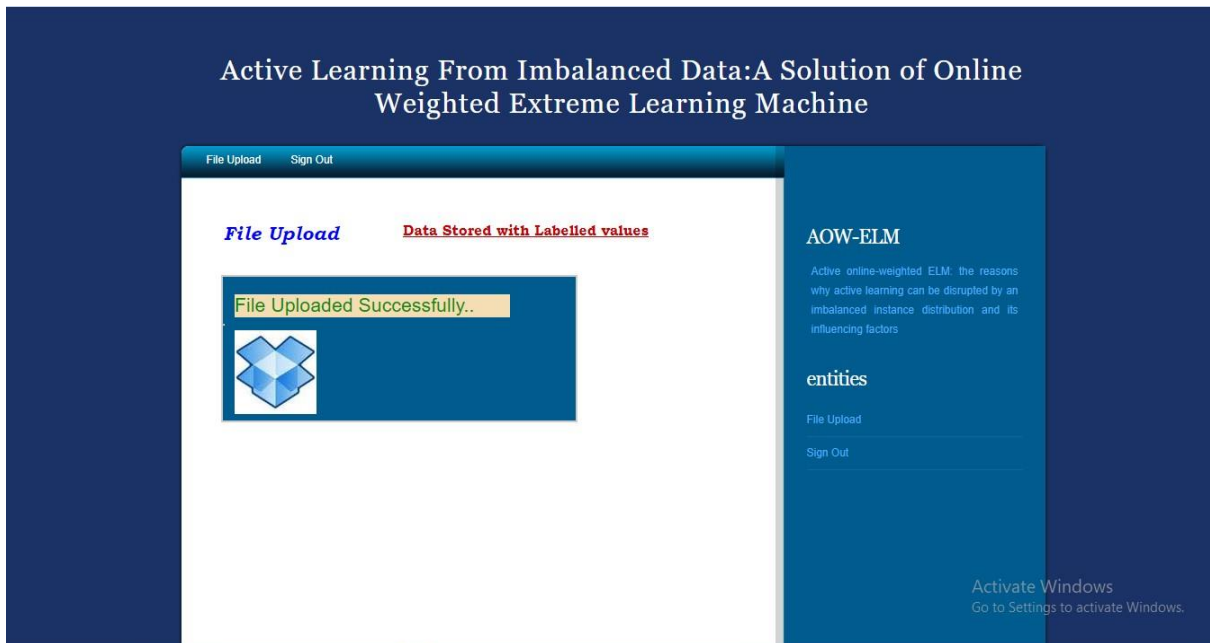


Figure 3 FILE UPLOAD STATUS

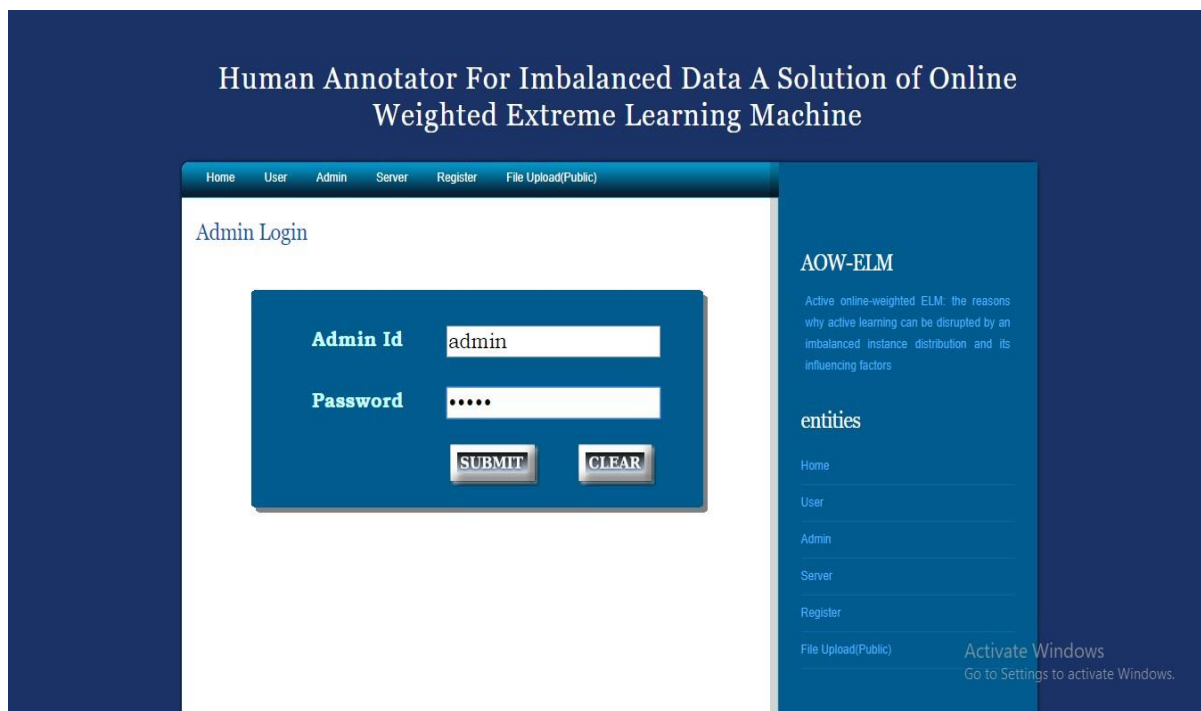


Figure 4 ADMIN LOGIN

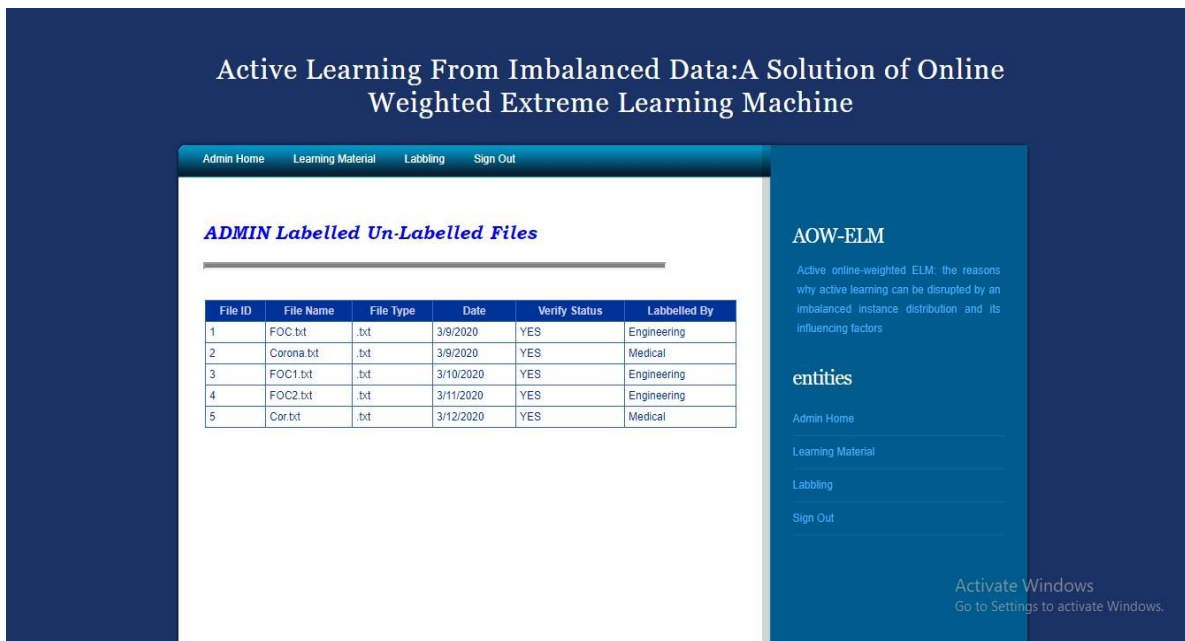


Figure 5 LABELLED AND UNLABELLED FILES IN ADMIN

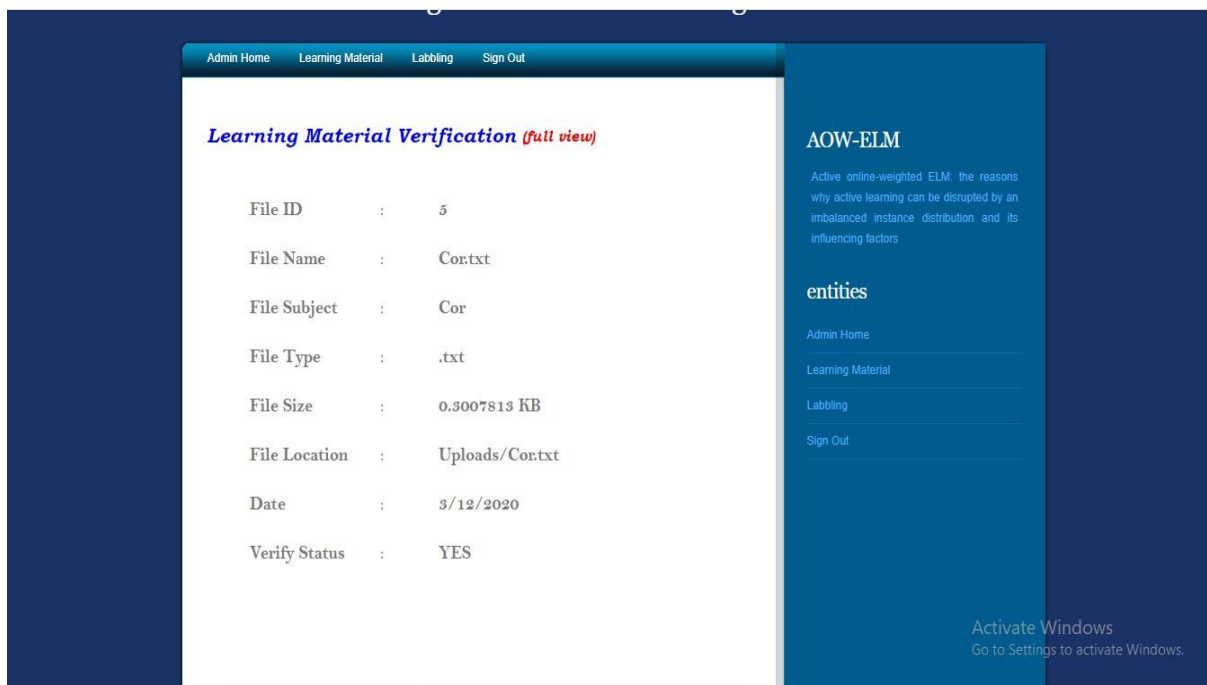


Figure 6 LEARNING MATERIAL VERIFICATION BY ADMIN

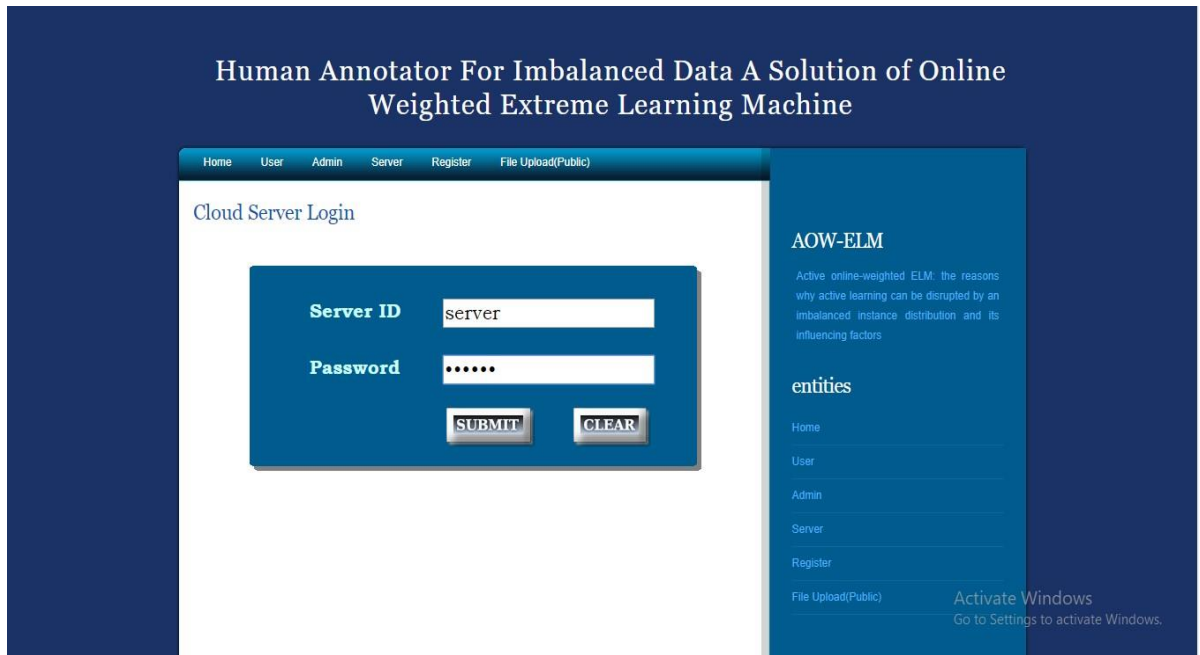


Figure 7 CLOUD SERVER LOGIN

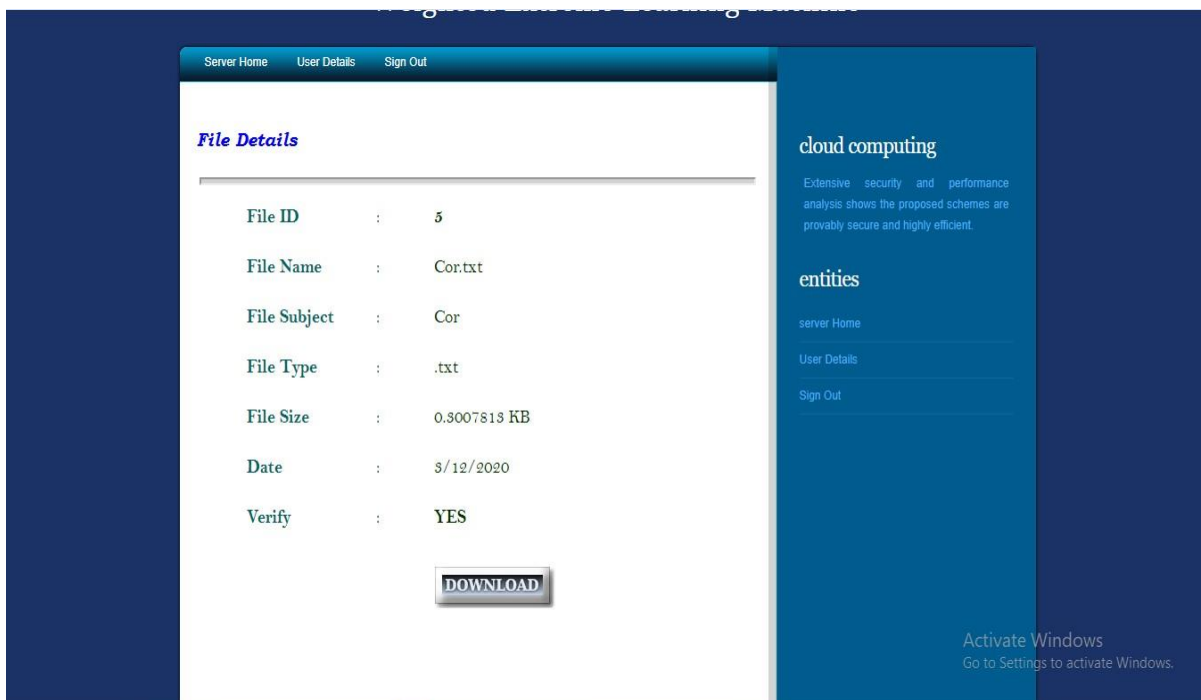


Figure 8 FILE DETAILS BY SERVER

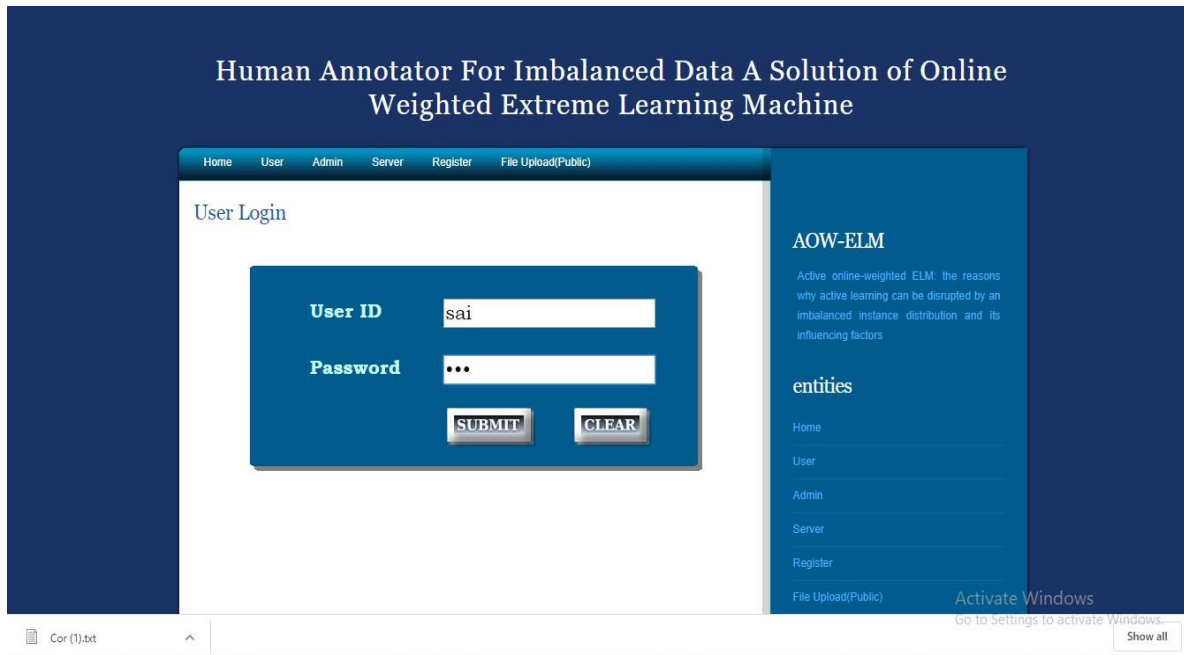


Figure 9 USER LOGIN

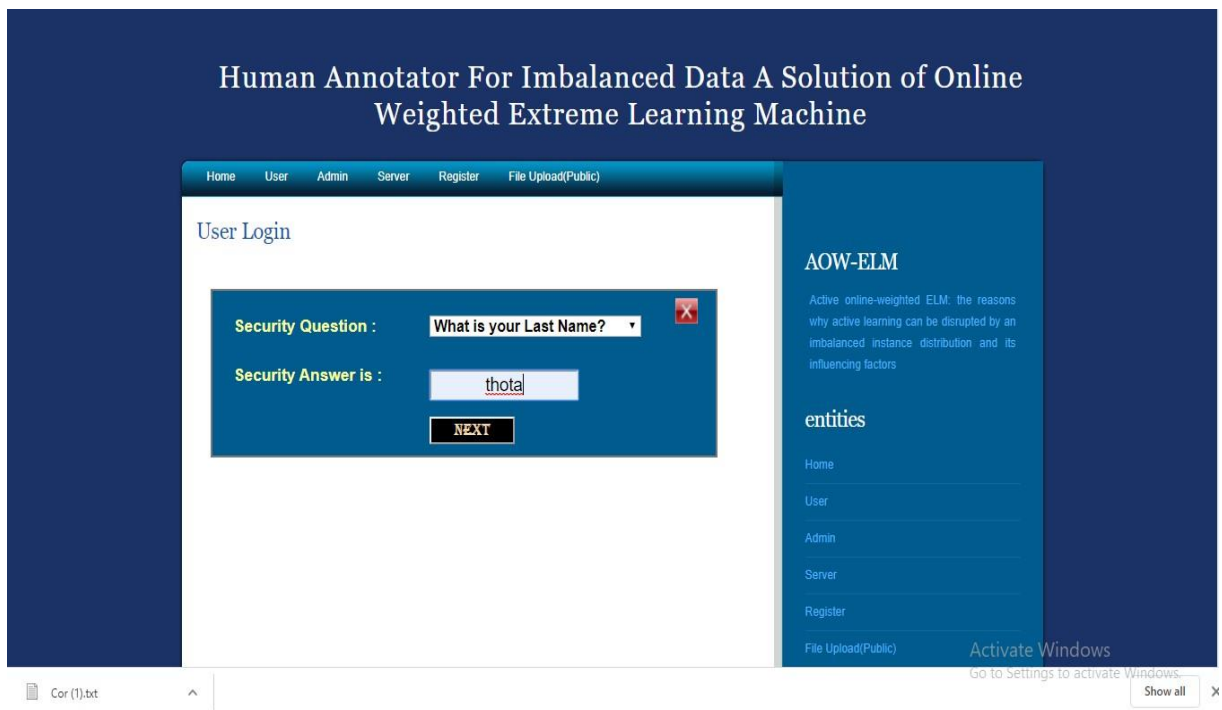


Figure 10 SECURITY QUESTION TO ACCESS DOWNLOADED FILE



Figure 11 SECURITY KEY FOR AUTHENTICATION

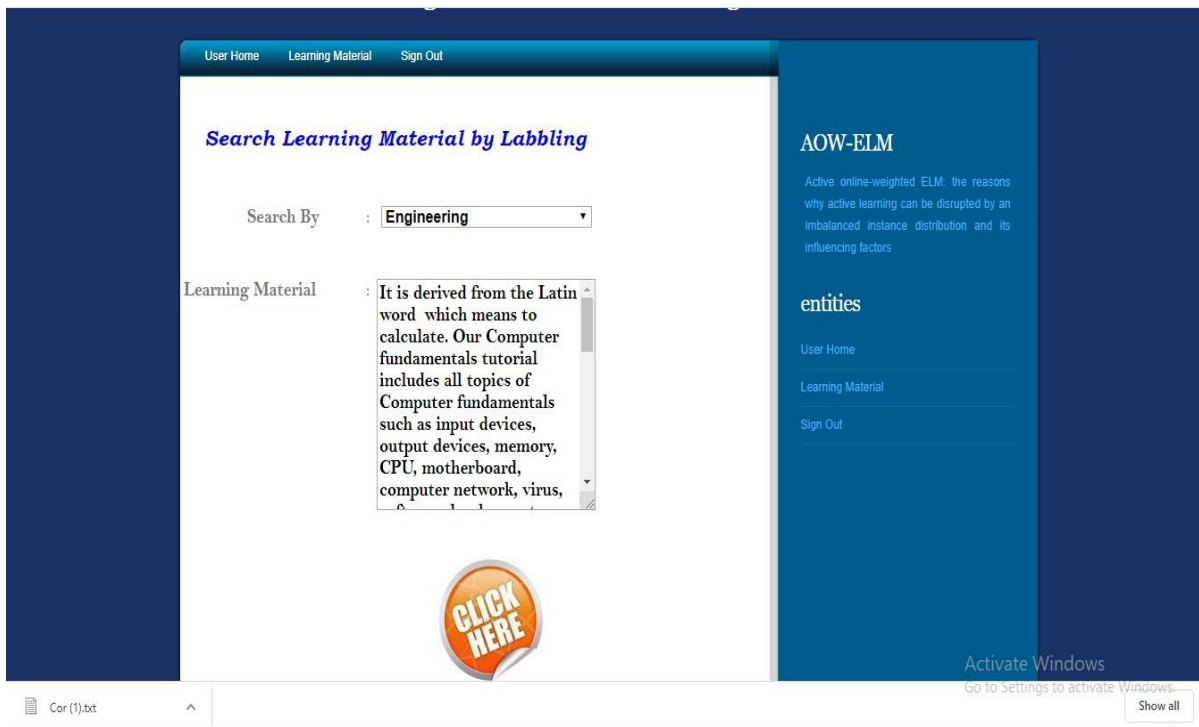


Figure 12 LABELLED MATERIAL FOR USER

C. PUBLICATION WITH PLAGARISM REPORT

**8th INTERNATIONAL CONFERENCE ON
“CONTEMPORARY ENGINEERING AND TECHNOLOGY –
2020”
(ICCET – 2020)**

March 14th- 15th, 2020

**ORGANIZED BY
ORGANIZATION OF SCIENCE AND INNOVATIVE
ENGINEERING & TECHNOLOGY (OSIET).**

Chennai, India.

Website: www.ijsiet.org

In association with

**PRINCE SHRI VENKATESHWARA PADMAVATHY
ENGINEERING COLLEGE
PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING
AND TECHNOLOGY
Medavakkam - Mambakkam Road, Ponmar, Chennai – 600 127**

ICCET 2020

***Proceeding of International Conference on Contemporary Engineering and
Technology. Journal: Journal of Engineering Science and Technology, Journal of current
Research in Engineering and Science.***

HUMAN ANNOTATOR FOR IMBALANCED DOSSIER

SAI KAVYA CHITTEM ^[1], SAI PRAVALLIKA THOTA ^[2], NIRMALRANI V ^[3]

[1][2] UG Student, Dept. of Computer Science, Sathyabama Institute of Science and Technology, Chennai, India

[3] Assistant professor, Dept. of Computer Science, Sathyabama Institute of Science and Technology, Chennai, India

ABSTRACT: - *Imbalanced dataset is pertinent fundamentally with regards to directed AI including at least two classes. Unevenness entails that the pile of info cynosures accessible for various classes are peerless. These informational indexes are extraordinary caisson for arrangement issue where the class conveyance isn't consonant. Normally, they are formed by two classes, the dominant part (negative) class and the minority (positive) class. Data about trial examines sets the Active learning is without a doubt compelling, however a few ongoing investigations have shown that dynamic learning is bombed when it is applied to information. In our venture Human Annotator will gather the information's from the open post and he will isolate named and unlabeled informational collections. Client need to enroll their subtleties and they can see their learning materials. Our thought is this human annotator will gather all the side effects of various maladies and he will isolate marked and unlabeled side effects of ailments and afterward coordinate that specific indications with coordinating side effects of individual infection by utilizing dynamic learning. The marked and unlabeled information's have been breaking down by human annotator and coordinated with named proper informational indexes are accomplished and it will be accessible for students.*

Keywords: *Active Learning, Human Annotator, Client, Learning Materials, Dataset.*

I INTRODUCTION

Numerous strategies have been grown as of late to battle imbalanced information grouping in machine learning. Information resampling strategies mean to adjusting information before applying standard order learning calculations. Two notable procedures can be used—stratified sampling [5], in which new information tests are produced between the minority information and their chose nearby. Tomek joins technique [6] and Wilson's altering [7] are under sampling strategies, in which a small amount of the greater part information are expelled. Adjusting information by resampling is useful in decreasing misclassification mistake, be that as it may, the viability isn't steady however issue subordinate [8]. Then again, new

machine was learning arrangement calculations or change of existing ones are likewise grown. In any case, cases may emerge in reality applications [9] when the misclassification cost network is hard to produce. To abridge the current dynamic learning calculations applied in the situation of unequal information disseminations, we found that they experience the ill effects of either low arrangement execution or high time-utilization issues.

Hence, right now, wish to propose a successful and effective calculation. The proposed calculation is named AOW ELM and it ought to be given in the selective sampling group mode dynamic learn situation with a vulnerability essentialness measure and E-Learning classifier.

Next, we conclude a proficient web-based learning method of WELM in principle and plan a powerful weight update rule [10]. At last, profiting by the possibility of the edge fatigue standard, we present a progressively adaptable and powerful early halting basis. In addition, we attempt to just talk about why dynamic learning [11] can be upset by slanted example conveyance, further researching the impact of three fundamental dispersion factors, including the class awkwardness proportion, class covering, and little disjunction.

In particular, we propose receiving the bunching strategies to beforehand choose the at first named seed set, and along these lines stay away from the missed group impact and cold beginning marvel however much as could reasonably be expected. Imbalanced information commonly alludes to order errands where the classes are not spoken to similarly. A large portion of this present reality characterization issues show some degree of class lopsidedness, which happens when there are not adequate examples of the information that compare to both of the class marks. Thusly, it is basic to pick the assessment metric of your model effectively. In the event that it isn't done, at that point you may wind up altering/enhancing a futile parameter. For instance, the datasets that manage false and non-fake exchanges [13], almost certainly, the quantity of deceitful exchanges as thinks about to the quantity of non-fake exchanges will be particularly less. What's more, this is the place the issue emerges. When the dataset has underrepresented information, the class conveyance begins slant.

What's more, besides, this can't at any rate ensure a proficient answer for your business issue. In most pessimistic scenarios, it may go to finish squanders with zero buildups to reuse. Managing imbalanced datasets incorporates different techniques, for example, improving

arrangement calculations or adjusting classes in the preparation information before giving the information as contribution to the AI [14]. Be that as it may, for inquire about purposes, in our task Human Annotator will gather the imbalanced data's, he will isolate named and unlabeled data's. Managing imbalanced datasets incorporates different procedures, for example, improving arrangement calculations or adjusting classes in the preparation information before giving the information as contribution to the machine learning. The last procedure is favored as it has more extensive application and adjustment. In addition, the time taken to improve is regularly higher than to create the necessary examples [12]. Yet, for inquire about purposes, in the venture Human Annotator will gather the imbalanced data's, and it will isolate marked and unlabeled data.

II RELATED WORK

[16] EL Machine was initially bounced for single layer feedforward systems anywhere info loads of a SLFN are haphazardly produced, and yield loads are prepared with bunch learning procedure of less squares. It has been demonstrated haphazardly shrouded neurons and tunable yield loads having all-inclusive estimation and astounding speculation execution [17]. All the more critically, ELM outflanks most existed learning calculations in preparing speed [18] and it has been broadly utilized in utilizations of face acknowledgment, picture handling and classification, power value arrangement, vitality product prospects record estimate, area fingerprinting strategy, protein succession order, and area grouping. **Yong Zhang et al [2]** Data streams with class awkwardness happen as a rule in numerous genuine applications.

Online consecutive learning is the one of the successful strategies for ordering information creek with class irregularity. This

paper come up with a double weighted online successive outrageous expert system technique to comprehend it. Based on online successive extraordinary one, the proposed dw-ELM strategy breaks down the circulation normal for information taking into account existence, and gives a versatile double weighting plan to tune the loads at both the time level and the space level. Broad trial assessments on 10 imbalanced datasets demonstrate that the technique beats a few looking at strategies as far as G-mean and F-measure measurements [19]. **Chengbo Lu et al[3]** for imbalanced information arrangement by consolidating casting a ballot technique into weight as outrageous learning machine (weighted ELM), these significant advances are associated with the proposed strategy: preparing weighted ELM classifiers, killing notusable orders to decide appropriate classifiers for casting a ballot, lastly deciding the characterization result dependent on greater part casting a ballot.

III EXISTING SYSTEM

Right now issue of dynamic learning in class lopsidedness situation of online weighted outrageous learning machine, they find that hurtfulness of slanted information dispersion is identified with various elements Datasets has low imbalanced proportion and little grouping covering, we need to allocate a huge edge for them to develop characterization models, perusers are propose to inspect the occurrence conveyance on the at first named set.

3.1 CONS of EXISTING SYSTEM

- Accurate data are not evaluated.
- Data sets are not appropriate.
- Multiple Dataset values are not labelled.

IV PROPOSED SYSTEM

In our proposed framework learning materials are posted by open clients any open people can give their information's .Learning materials will be seen by Human Annotator and he have to isolate named and unlabeled information's and need to assign a mark or need to proceed with coordinating name ,when the information's are named it will be accessible for students. Confirmation is about check of your certifications, for example, Username/User ID and secret key to confirm your character. The learning material will be transfer by Public post, material can be transferred by named or unlabeled material and all the materials will be put away in the server. Human Annotator will gather all the named and unlabeled learning materials named materials will be noticeable in Active learning and unlabeled information will be marked by human annotator and that material additionally accessible for Active learning. We proposed an AOW-ELM; In particular, we additionally demonstrated that there is an inexact mapping connection between the genuine yields of ELM also probabilities in the Bayes classifier. The material will be accessible for the client when all the information's are have been named and client can pick the taking in materials from the accessible materials and can see the definite learning material and materials can be downloaded for their utilization.

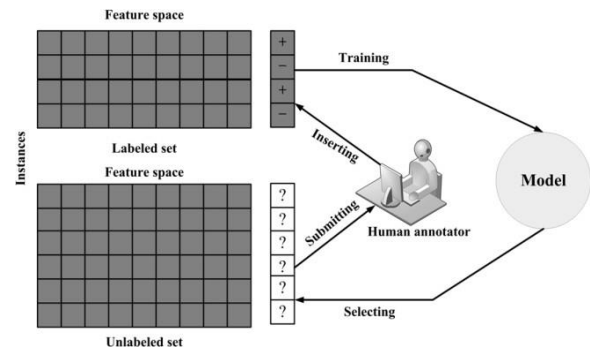


FIG 1 SYSTEM ARCHITECTURE

4.1 PROS of PROSPECTED SYSTEM

- No loss of data.
- It's easy to actualize and decipher.
- It can help improve the runtime of the model and take care of the memory issues by decreasing the quantity of preparing information tests when the preparation informational index is gigantic.

V MODULES DESCRIPTION

5.1 AUTHENTICATION AND AUTHORIZATION

Confirmation is about check of your accreditations, The framework investigation, regardless of whether you are utilizing your qualifications or not. As a rule, confirmation is finished with a username and secret word, in spite of the fact that there are different approaches to be verified. Be that as it may, the approval procedure of this is to give access as endorsement to the client. So that after the endorsement the one can ready to access by confirming your privileges.

5.2 MATERIAL UPLOAD

The learning material will be transfer by Public post, material can be transferred by marked or unlabeled material and all the materials will be put away in the server, Human Annotator will gather all the named and unlabeled learning materials named materials will be unmistakable in Active learning and unlabeled information will be named by human annotator and that material likewise accessible for Active learning.

5.3 ACTIVE LEARNING

The advantages of dynamic learning is in reducing both in the heaviness of a human

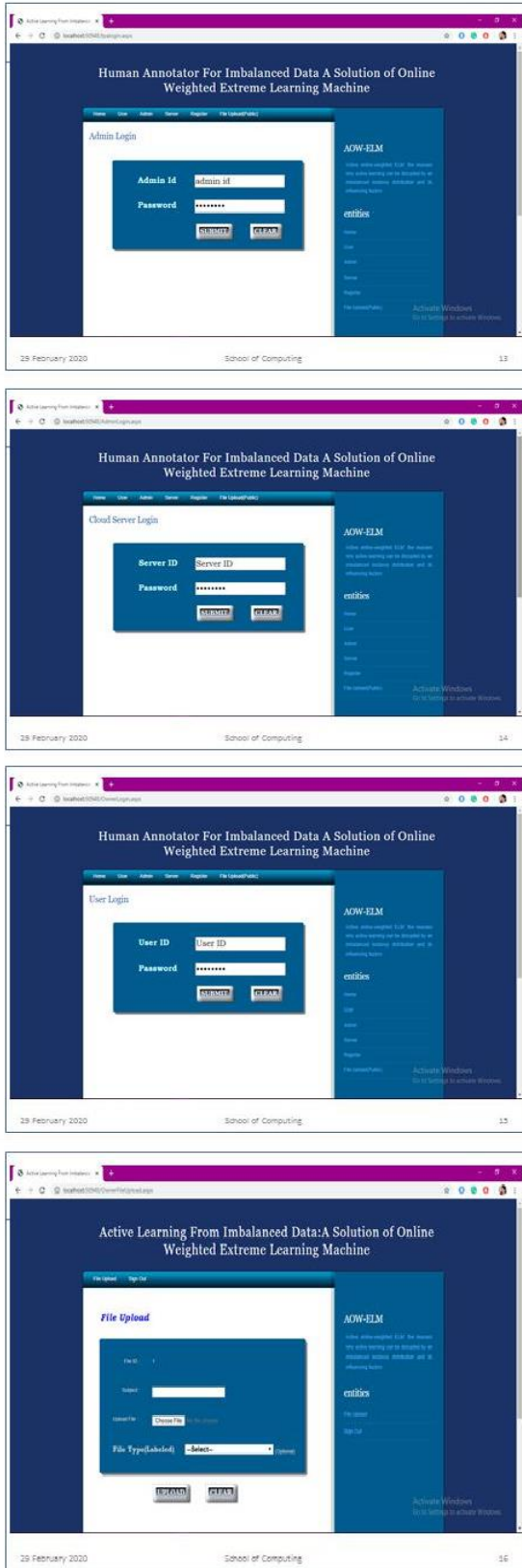
authorities and the eccentrics of planning events anyway getting a request model that passes on preferred or for all intents and purposes indistinguishable execution over the model with denoting all events. It is outstanding that dynamic learning can at the same time the idea of the gathering the multifaceted design of getting ready models. Regardless, a couple past examinations have demonstrated that the introduction of dynamic learning is viably irritated with an imbalanced data scattering. Some present imbalanced powerful adopting in strategies also experience the evil impacts of either low execution or high time usage.

5.4 AL WITH EXTREME LEARNING MACHINE

We intended a learning computation reliant on ELM and given a name it as AOW-E-LearningMachine. We originate that the genuine yields of ELM can cast the powerlessness lay of events, i.e., their gathering determinations. Peculiarly, we moreover exhibited that there is a harsh; mapping association between the genuine yields of ELM and Bayes classifier.

VI RESULT AND DISCUSSION





VII CONCLUSION

Right now, improved calculation contracted for tackling double imbalanced arrangement issues IW-ELM has been proposed. We investigate the issue of dynamic learning in class irregularity situation. We find that the destructiveness of slanted information conveyance is identified with various factors, and can be viewed as a blend of these components. Various leveled grouping can be adequately used to beforehand remove beginning agent examples goesto a seed set where it is to domicile the abeyant mislaid bunch impact and bleak beginning wonder. The examination between the proposed machine calculation and some other benchmark calculations demonstrates that AOW-ELM was a successful technique to address the issue of dynamic learning in a class irregularity situation. Reenactment results showed that IWELM accomplishes higher precision contrasted with other ELM based calculations.

REFERENCES

- [1] B. Settles, -Active learning literature survey, Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. 1648, Jan. 2010.
- [2] Yong Zhanga,b, WenzheLiua, Xuezhen Rena and Yonggong Rena, -Dual weighted extreme learning machine for imbalanced data stream classification Journal of Intelligent & Fuzzy Systems 33 (2017) 1143–1154 DOI:10.3233/JIFS-16724 IOS Press 1143
- [3] Chengbo Lu · Haifeng Ke · Gaoyan Zhang · Ying Mei · Huihui Xu3 , -An improved weighted extreme learning machine for imbalanced data classification, Received: 2 January 2017 / Accepted: 20 June 2017 © Springer-Verlag GmbH Germany 2017

- [4] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Stream-based active learning for sentiment analysis in the financial domain," *Inf. Sci.*, vol. 285, pp. 181–203, Nov. 2014.
- [5] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Batch mode active learning and its application to medical image classification," in *Proc. ICML*, 2006, pp. 417–424.
- [6] D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Proc. ICML*, 1994, pp. 148–156.
- [7] M. Lindenbaum, S. Markovitch, and D. Rusakov, "Selective sampling for nearest neighbor classifiers," *Mach. Learn.*, vol. 54, no. 2, pp. 125–152, 2004.
- [8] S. Dasgupta and D. Hsu, "Hierarchical sampling for active learning," in *Proc. ICML*, 2008, pp. 208–215.
- [9] R. Wang, S. Kwong, and D. Chen, "Inconsistency-based active learning for support vector machines," *Pattern Recognit.*, vol. 45, no. 10, pp. 3751–3767, 2012.
- [10] S. C. H. Hoi, R. Jin, and M. R. Lyu, "Large-scale text categorization by batch mode active learning," in *Proc. WWW*, 2006, pp. 633–642.
- [11] N. Roy and A. McCallum, "Toward optimal active learning through Monte Carlo estimation of error reduction," in *Proc. ICML*, 2001, pp. 441–448.
- [12] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *Proc. ACM SIGIR*, 1994, pp. 3–12.
- [13] J. He and J. G. Carbonell, "Nearest-neighbor-based active learning for rare category detection," in *Proc. NIPS*, 2008, p. 633–640.
- [14] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," in *Proc. NIPS*, 1995, pp. 231–238.
- [15] K. Fukumizu, "Statistical active learning in multilayer perceptrons," *IEEE Trans. Neural Netw.*, vol. 11, no. 1, pp. 17–26, Jan. 2000.
- [16] Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3):489–501
- [17] Zhang R, Lan Y, Huang GB, Xu ZB (2012) Universal approximation of extreme learning machine with adaptive growth of hidden nodes. *IEEE Trans Neural Netw Learn Syst* 23:365–371
- [18] Cao J, Hao J, Lai X et al (2016) Ensemble extreme learning machine and sparse representation classification. *J Franklin Inst* 353(17):4526–4541
- [19] Li X, Mao W, Jiang W et al (2016) Extreme learning machine via free sparse transfer representation optimization. *Memetic Comput* 8(2):85–95
- [20] Cheng X, Liu H, Xu X et al (2016) Denoising deep extreme learning machine for sparse representation. *Memetic Comput*. doi:10.1007/s12293-016-0185-2

PLAGARISM REPORT

Turnitin Originality Report

Processed on: 10-Mar-2020 11:22 PM CDT

ID: 1273450006
Word Count: 2419
Submitted: 1

07 By 07 07

Similarity Index	Similarity by Source
	Internet Sources: 4%
6%	Publications: 4%
	Student Papers: 2%

1% match (Internet from 01-Feb-2020)

<https://link.springer.com/article/10.1007%2Fs12293-017-0236-3>

1% match (Internet from 29-Oct-2019) h

<https://ijsiet.org/vol12018/143.pdf>

1% match (Internet from 17-Jan-2020)

<https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs16724>

1% match (publications)

[Hualong Yu, Xibei Yang, Shang Zheng, Changyin Sun. "Active Learning From Imbalanced Data: A Solution of Online Weighted Extreme Learning Machine", IEEE Transactions on Neural Networks and Learning Systems, 2019](#)

1% match (student papers from 11-Feb-2017)

[Submitted to Shri Guru Gobind Singhji Institute of Engineering and Technology on 2017-02-11](#)

1% match (publications)

[Chengbo Lu, Haifeng Ke, Gaoyan Zhang, Ying Mei, Huihui Xu. "An improved weighted extreme learning machine for imbalanced data classification", Memetic Computing, 2017](#)