

Availability and Load Balancing in Cloud Computing

Nadia Tabassum¹, Shahida Mujeed².

Department of Computer Science, Virtual University of Pakistan, 54000, Pakistan^{1,2}

*Corresponding Author: *Shahida Mujeed Email: sabaawan2903@gmail.com

Received: XX Month 202X; Accepted: XX Month 202X

Abstract

Cloud computing is new model that permit to the clients, associations, to buy the necessary administrations as indicated by our requirements. It is used to upload their data and retrieve data according to the needs over the internet. This model offer several services like to store data, easy and convenient web services, etc. In this era Cloud Application is developed the different services some are Platform-as-a-Service (Paas), Software-as-a-Service (Saas), Infrastructures-as-a-Service (Iaas), and Cloud computing improve their services day by client also demand reliable and new services for efficiency and reliability. Load Balancing be major provocation in cloud computing. Technique called load balancing is used to distribute load of the data over the cloud network. It is also used to minimize the resource usage.

Keywords: Cloud computing; QoS parameter; cloud service providers; ranking; prediction.

1. Introduction

A tool called Cloud computing permits and is convenient according to the user demand. It is very famous around the world last some years they provide flexibility also easy to access data retrieval and files over the internet also provide facility to send and receive large data spread around the world easily through cloud. Cloud computing show a continuous growth of the business due to cause

of less expenses. Users increase day by day we use some technique to handle the large data required different methods to optimize and streaming operations that provides best and satisfactory levels of performance for the user. [1 (Mortada, 2020)].

0.1 Cloud Service Models

Service can be described as the endpoint of a connection; it should be well-defined, self-explained, and independent of context.

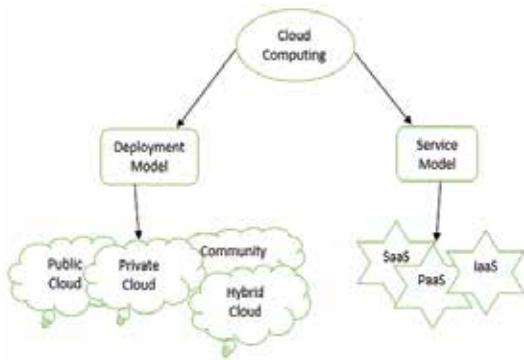


Figure 1. Cloud Deployment and service model

0.1.1 Software as a service (SaaS).

In cloud software as a service also called software on the demand of the user or customer. This service can be available on any part of the world by subscription or consumer based. Some examples are Drop Box, Google Apps, Microsoft office 365, big commerce etc. [2 (Geetha, 2017, August)]

0.1.2 Infrastructure as a service (IaaS).

In cloud infrastructure as a service (pay as you go for the service like storage, networking and virtualization) has the opportunity to grasp the basic in other worlds we say root infrastructure of providing the network operating systems, data storage, & many other basic cloud computing materials where user can run and deploy the software. It is time saving and have better flexibility also provide Resource Management and Remote Access. E.g., are, Amazon Web Services (AWS), Microsoft Azure and Google Compute Engine (GCE) etc. [(Chen, 2018)2]

0.1.3 Platform as a service (PaaS)

In Cloud platform as a service offering many software and hardware tools available over the internet, APIs and allow impulsively deployed on virtual Machine Infrastructure. In this era

there is no generic methodology are followed. In PaaS multiple user adaptable and ideal for the business point of view where multiple experienced developers are working on the single project. Some examples are Force.com, Microsoft Azure, and AWS etc.

1.2 Deployment Models

Four deployment models are currently used in cloud computing.

1.2.1 Public Cloud

Public Cloud model is the most extensively used cloud service. Public Cloud type is an admired preference for web applications, file sharing and insensitive data storage. Service supplier owns and manipulate the hardware's pass needs to run for a public cloud. Public cloud play a very essential role in testing and development.

1.2.2 Private Cloud

Private cloud is basically used for a specific organization. Private cloud organization control the system and manages it in a compact fashion. While a 3rd party can host a Private cloud server, mostly the organizations select and support the hardware in their own place data center. Only authorized person can access data in private cloud. It has a full owner control.

1.2.3 Hybrid Cloud

It is a combination of two or more infrastructure. Every infrastructure in this model have a separate system, but they all are the part of the same architecture. It has high set-up flexibility, low cost.

1.2.4 Community Cloud

Model called Community cloud, work same as a public cloud but the main difference is that

community cloud allow a specific access group of persons with shared interest and use cases. This architecture can be hosted according to own rules, or 3rd party provider and a spy

organization. In cloud community, all organization have the same policies for security, government issues and application types.

Comparison:

Types	Private Cloud	Public Cloud	Community Cloud	Hybrid Cloud
Description	Used for specific organizations. Private cloud is build up with existing private infrastructure.	Services available for public users.	Different type of cloud are integrated together to meet a common or particular need for some organization.	Hybrid cloud is a distributed heterogeneous distributed system.
Scalability	Limited	Very high	Limited	Very high
Reliability	Very high	Moderate	Very high	Medium to high
Security	High class security	Totally depend on service model	Secure	secure
Performance	Good	Low to medium	Very good	Good
Cost	High cost	Cheaper	Costly	Costly
Examples	VMWare, Microsoft	Amazon EC2, Google App Engine	SolaS Community cloud	IBM, HP, V Cloud

2. Cloud Architecture

Cloud computing is divided into two main parts: the front end and the back end. It gives you the apps and interfaces you'll need for your Cloud-Based service. It is based on web browsers such as Google Chrome and Internet Explorer, which are client-side programmers. Virtualization, data storage, and other software and hardware components are included in cloud infrastructure. Cloud computing includes a graphical user interface (GUI) that

may be used to control views and execute applications on the front end. It then examines the economics of cloud computing in terms of lowering costs, converting capital expenditures into operational expenses, boosting quantity demand, improving accuracy, and lowering latency.[(Asghari, 2021)] Some advantages are it solves latency issues, Reduce costs, its helps businesses, Flexibility features, High security, Encourages remote access, Auto services.

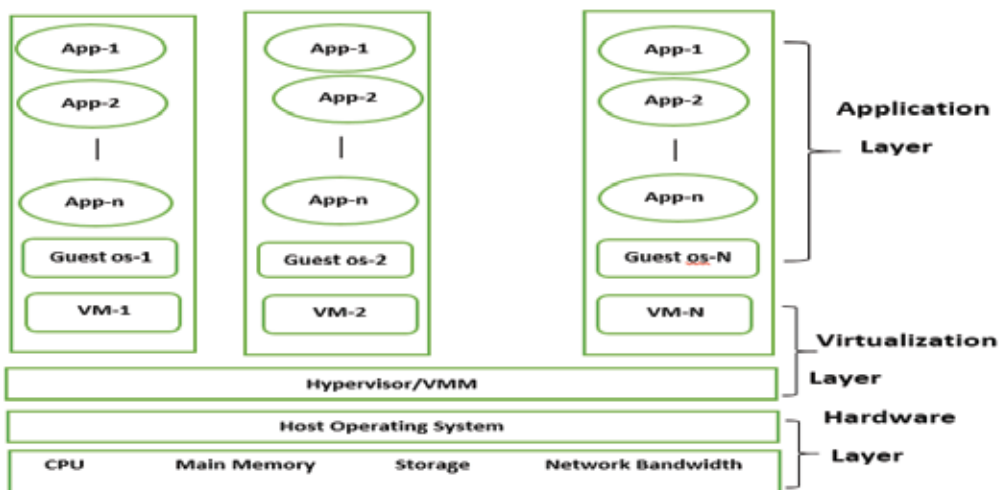


Figure 2. Single Host Architecture

3 Literature Review

After study, the different algorithms and techniques and cloud software's I am able to write something about the cloud computing here some point of views. Literature deal with Heuristic Algorithm & Nature Inspired Algorithm it's the part analysis the advantages for cloud analyst. Li and Wu (2019) presented the Load Balancing Ant Colony Optimization (LBACO) algorithms & find best offers that are finding in Load Balancing coupled with best make-span. [4 (Al Nuaimi, 2012)]

In other hand, the best simulation results given to us when the algorithm provided 30 % boost the execution time, waiting time & make-span compared to earlier algorithm with highest resource utilization and minimum overhead. [5 (Arulkumar, 2021)]

We examine about the difficulties of the entrance in an application improvement Methodology For programmed advancement of cloud based business application this is fundamentally intended for the non-it client resembles' partners and business specialist for application development. We have find the administrations for work process, script produced likewise test the programmed organization deployment process measure [6 (nzinger, 2014)].

Despite the various previous research organizations in the field of Cloud Computing, numerous questions remain associated with duty shifting in applications based on cloud and, more explicitly, in the IaaS model. IaaS is one of the innovation models that looks at the back end, where management of servers, data

centres, and virtual machines takes place. In such setups, Providers of Cloud Service should provide high direct help delivery, avoiding situations like these. For example, having being overburdened or under-stacked, as this will result in a longer performance time or machine collapse, and so on Errand booking significantly adds to stack adjusting, and planning assignments closely sticks to the Service Level Agreement (SLA) requirements, a document provided by cloud engineers to clients.[7 (Panda S. K., 2019)]

In the LB computation, significant SLA limits, such as Deadline, are taken into account. Considering the Quality of Services task specifics, the requirement for VMs, and asset identification, the suggested computation is predicted to improve assets and improve Load Balancing. The suggested LB Algorithms address the concerns raised as well as the ebb and flow research hole caused by the writing's conclusion. The suggested LB computation yielded a typical of 78% asset utilization when compared to the genuine Dynamic LBA Algorithm. It likewise accomplishes great execution regarding less Execution time and Make span.[8 (; Yoo, 2011)]

The searcher will try to further optimize cloud resources and improve cloud-based application performance, such as for as many SLA criteria as possible. For example, for good outcomes, the algorithm will be assessed based on the number of infractions and migration count. In addition, the method will be compared against other algorithms in the literature. [9 (Mallikarjuna, 2019)]

One of the major struggling in cloud computing is Load balancing. This paper

focus mostly on how load balancing affects new technologies that rely on cloud computing. Furthermore, effective load balancing on the cloud using IoT, Big Data, & Self-learning systems can be used to achieve the present ways of trustworthy and dependable cloud computing. The effectiveness and value of the DBLA and NDLBA algorithms for next-generation cloud computing have been demonstrated. The main goal of this research is to increase overall achievement and to maximize revitalization for task assignment on virtual machines. This paper explain the distinct static and dynamic load balancing algorithms & its types. In the coming up, the need to design fully autonomic new Dynamic load balancing algorithms empower good resource utilization, minimum make span [88], improvement in the degree of imbalance, effective task migrations, and minimum make span of time for next generation cloud computing.[10 (Jena, 2020)]

4 Characteristics

Here we discuss some characteristics about the cloud computing.

4.1.1 On Demand Self services:

Cloud services deliver services without requiring the user or the service provider to interact. Amazon Web Services, IBM, and Microsoft are examples of cloud service providers that deliver services based on customer requests such as self-services.

4.1.2 Broad Network Access:

Cloud potential are at hand on the network and obtained the standard mechanized that provide and promote the different users who used

laptops, mobiles etc.

4.1.3 Rapid Elasticity:

Rapid elasticity have the ability can be elastically the girding and let out, in some situations automatically, to scale rapidly external and internal with some demands. Here some applications adding or removing the nodes, resources and services. [11 (Ibrahim, 2021)]

4.1.4 Measured Services:

In this case automatically optimized and control the recourse usage grip a metering capability at some level of abstraction in an appropriate way. The recourse usage can be monitor, reported and control & also providing some glassiness for consumer as well as the provider for the provide on-self services.

4.1.5 Resource Pooling:

The service provider providing the computing recourse pooled together and to serve multiple customer, with different customer behavior like some physical & virtual resources randomly assigned and re-assigned according to the user or customer demands. In this case customer or not bounded “geographically” but they may have some specific location at higher level query. Some examples are storage, processing, network bandwidth and memory. [12 (Lyu, 2020)]

4.2 Cloud Infrastructure

1. The cloud services are accessed through web services over the Internet. Cloud service providers offer versatile services such as computing services, storage services, and content delivery networks [13 (Khan, 2020)]. A single service can be offered in multiple regions. The main module of the proposed model is the QoS ranking module, which is further divided into three sub-modules:

- Similarity analysis module
- QoS parameter module
- Training data module

2. Similarity analysis can be used to identify cloud services of the same type according to the service model, e.g., to differentiate between software-as-a-service or platform-as-a-service [14 (Wagle, 2015)]. QoS parameters can be used to rank the cloud services according to their respective service category. The QoS parameters used are

- Security feathers
- Availability of cloud services
- Downtime
- Outage
- Response time
- Price
- Trust.

Cloud services are compared with the standard cloud benchmark [15 (Kumar, 2019)].

4.3 Cloud Feathers Extractor

All of the cloud services are listed and given

to the input layer. First, the services are selected, that is, the cloud services, storage services, and computation services. Next, the input data are given to the next layer for further extraction of the QoS parameter [16 (Ghahramani, 2017)]. Further sampling is done based on QoS parameters.

5 Load Balancing

Load balancing is very important and play a vital role in clouds environment they optimized the resource in VMs in Cloud computing. They also deals the equality of recourse and load distribution of work in an efficient way in the resource utilization. The systematic balanced of work load take higher user gratification & good resources usage. Load balancing decrease the delay sending and receiving data as well as the intercept the overburdened the situation in the nodes that effected the Qos in the data centers of the cloud [17 (Ala'Anzy, 2019)].

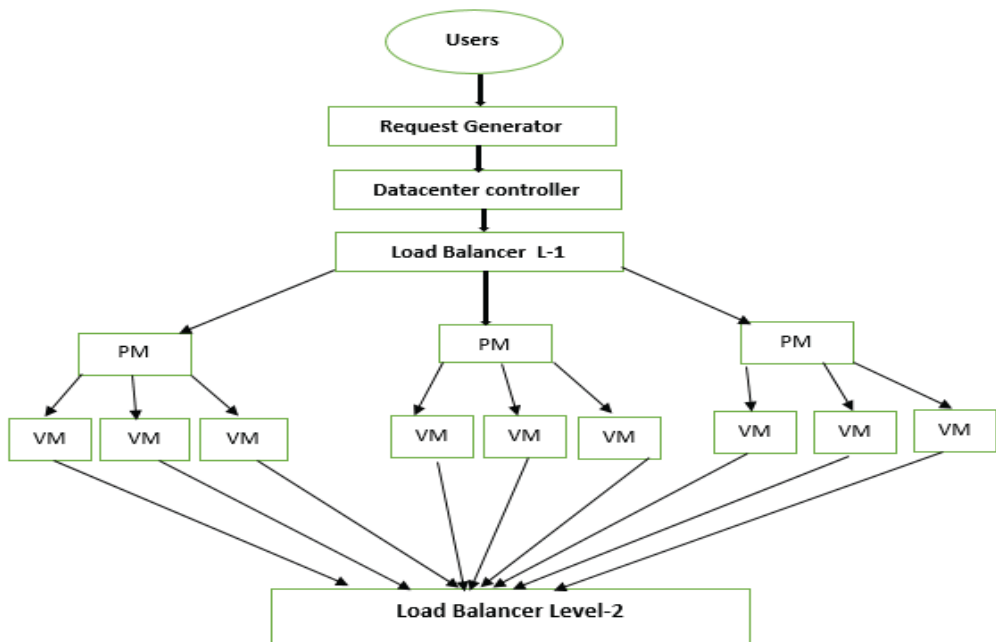


Figure 3 Load Balancing Architecture

As shown in this diagram, two types or designs of load balancing design models are clarified in unbalanced clouds for load scheduling. Gupta et al. propose a different architecture. Machine Virtual This approach abstracts the VM management and VM screen. The primary level load balancing is done by The Physical Machine (PM) level, whereas the Virtual Machine (VM) level performs the sub sequential level. According to the outcome, there are two sets of experiments here. [18 (Afzal, 2019)

1. Intra VM task migration
2. Inter VM task migration

Client requests that need calculating assets for their implementations are generated by the solicitation generator. The server farm regulator is in charge of objective management. The heap balancer determines which VM to assign to a certain client job. The major source of stress by distributing responsibilities among its own linked Virtual Machines, the Balancer modifies the provided Workload on particular Physical Machines. The burden balancer at the next level balances responsibilities among multiple Virtual Machines and Physical Machines. Exercises involving load adjustment the distributed computing responsibility is established by organizing and dispensing tasks to virtual machines based on their requirements. The actions that make up the load balancing process are as follows.

- User task requirements recognition
- resource details of a VM recognition
- Scheduling tasks
- Allocation of resources
- Migration

6 Types of cloud computing:

6.1.1 Round Robin:

The Round Robin strategy depends on a

revolution framework to sort network and application traffic. An inbound solicitation is designated to the most readily accessible worker, and afterward the worker is knock the lower part of the line. This strategy is especially helpful when working with workers of equivalent worth.

6.1.2 IP Hash:

In this direct load balancing method, the customer's IP address essentially figures out which worker gets its solicitation.

6.1.3 Least connection:

As its name expresses, the least association strategy guides traffic to whichever workers has minimal measure of dynamic associations. This is useful during heavy traffic periods, as it keeps up even conveyance among every accessible worker.

6.1.4 Least Response Time:

The least reaction time strategy guides traffic to the worker with minimal measure of dynamic associations and most reduced normal reaction time.

6.1.5 Least Bandwidth:

This application load balancer strategy estimates traffic in megabits (Mbps) each second, sending customer solicitations to the worker with the least Mbps of traffic.

7 Load Balancing Algorithm

When a client enters some information into the computer her data produce a task & reach the information center controller. Data center controller at the point apportions this to the load balancer at the starting steps. At that point they asked for the routed to abundant types of

Physical Machines. Inside the physical machines, there exists numerous of various machines which are made as per wants load balancer is transferred in such way to create the stack to be balanced, & smooth exchange is accomplished.

The most objective of any cloud based load balancing algorithm is to supply the administrations to the client without any delay, and smooth exchanges are anticipated b/w the client and the cloud benefits supplier . The cloud service provider's will be in contract with the conclusion client within the form of SLA.[]

Figure 5 shows how to put this underused algorithmic notion into practise. We'll need a few tools to help us with this. We can code in a variety of languages, including Java, PHP, Python, C#, and others. This extension has been updated using PHP. Encouragement, we'll need a text editor, such as Net beans, Eclipse, or Dreamweaver. however, if we want any support mainly back-end for data bases, we may use Xampp, Wamp, SQL Server, MS Access, and so on. Cloud users can connect to the internet via a wired or remote network.

Acquiring IP addresses of the cloud nodes is the basic principle underlying the network. To initiate, this application will ask the client for a request that may be addressed using any of the assets available. When a client submits a request, the request is sent to the Ace server for load balancing. For the client, this handle is used in the foundation. Currently, a stack balancing server sends requests to various levels of the framework and then to the final product. With the aid of the flow chart in the accompanying Fig, describes the load balanc-

ing algorithm in further detail.

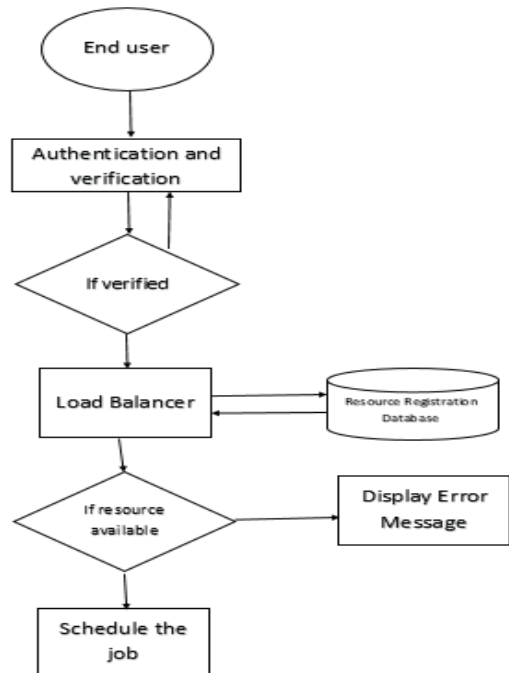


Figure 4. Flow chart of load balancing Algorithm

7 Proposed Methodology

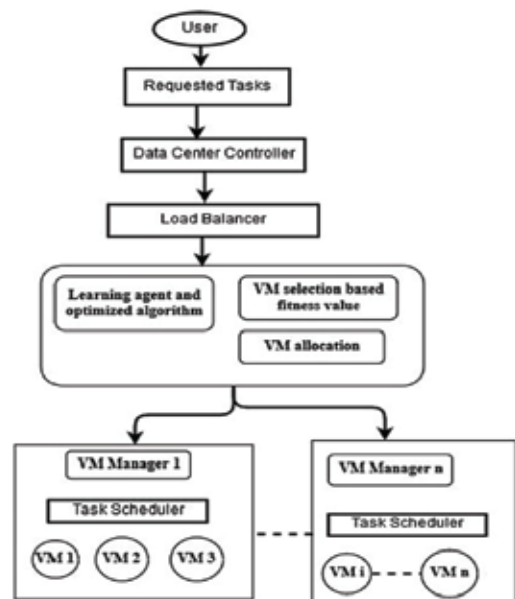


Figure 5 Proposed Methodology

In the above Fig.5. Explain about the propped methodology with a picture for excellent understanding.

Utilizing practical resources whose size or capacity of tasks fluctuate during run time in an irregular fashion, load balancing aims to increase the compilation time of a programme or job. When there is little variation in the load on the virtual machines, static load balancing techniques will be expensive. Because of the loads' unpredictable movement during run time, the static load balancing algorithm will thus not function properly. As loads change as applications run, dynamic load balancing is more advantageous than static load balancing and should take stack data and maintenance into account. Due to the network's rapid development and the demand for resources during operation, dynamical methods and techniques are very important and effective in distributing the load among the various resources. Our hybrid meta-heuristic algorithm is a dynamic approach to load balancing and sets the task's requirement in the queue of VMs that are waiting. This system has carried out user requests for the task. (Sambit Kumar Mishra a, 2020). [(Gundu, 2020)]

The long-term data centres and higher performance for external services are the goals of this research.

Data centres are typically located away from end users. Assignments should be shifted from heavily loaded VMs and distributed to under-loaded VMs when multiple tasks are simultaneously assigned to a certain VM and there are still available VMs nearby on the network. As a result, the multiple tasks or assignments can be distributed among all VMs with a combined first priority, which reduces the task's waiting time while also increasing VM throughput and enabling load balancing at the VM level. The load balancing framework concept is depicted

in the above figure, where the client requests that the task be carried out on the host. The task given to the load balancer by the board is the data center's responsibility. In order to distribute the assignment to the VMM, the load balancer would be chosen. Verified and active VMM demanded a lot of resources, and the host's resources had to be available. If obtaining VMs is insufficient, it must generate additional VMs in response to task requests. In this way, load balancing will be carried out based on the best value of the VMs since each host can only support a certain number of VMs. [(cloud)]

8. Classification of Load Balancing Algorithms:



Figure 6 Classification of load balancing Algorithm.

Static Strategies:

The Static Load in Cloud Computing Strategies is usually categorized into two categories. Physical machines are available at the beginning of both the initial errand restaurant and the subsequent assignment. After each task has been booked, the asset will be completed. OLB, MET, MCT, GA, Switch Algorithm,

TABU, A*, MIN MIN, MIN-MAX, and other heuristics in static technique include: OLB, MET, MCT, GA, Switch Algorithm, TABU, A*, MIN MIN, MIN-MAX, and so on. [(Gundu, 2020)]

Dynamic strategies:

This is an important cloud computing methodology. During the run period, they also load distributed the heap among thy Physical Machine (PM). Off-line mode (Batch mode) and On-line mode are two ways to represent these heuristics-based powerful calculations used for load adjustment. The job is allocated exactly at certain preset minutes in Off-line mode heuristics. It is used to determine the true cause for the execution of a large number of jobs. MAX-MIN, MIN-MIN, and Suffrage Algo are some of the newly added heuristics for Off-line modes. When a client interest (task) enters the scheduler in On-line mode

(prompt mode), it is organized onto a figuring center point. Each attempt is reserved for a single time, and the booking outcome remains unchanged. OLB, MET, MCT, and SA are some of the heuristics offered for online modes. [(Gundu, 2020)]

9. Result

I try various models : “time Series Models”-like WayeNet, LSTM, ResNet-1d, FCN-1d. Furthermore, I saw that the first preparing informational index is unequal - there is not many information with an opportunity to disappointment of over 8 seconds. Hence, while making my own dataset for train the model, I utilized information from 2, 7, 14 "long" shudders "more" than others, and utilized fourth tremor for approval. This improved the forecast for longer occasions to disappointment. show

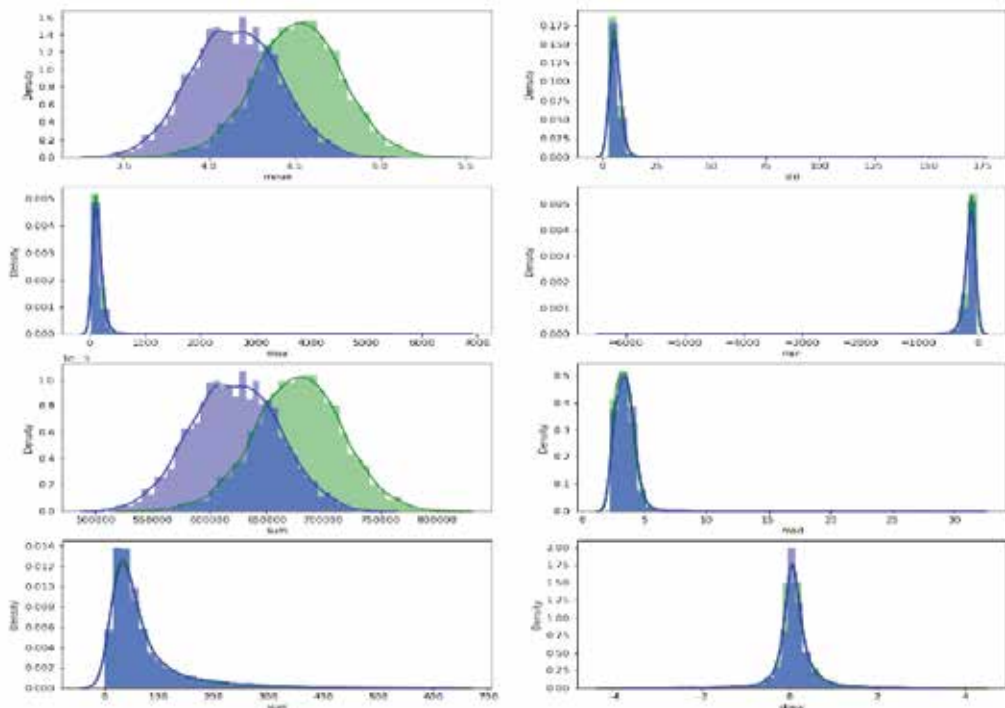


Figure 7

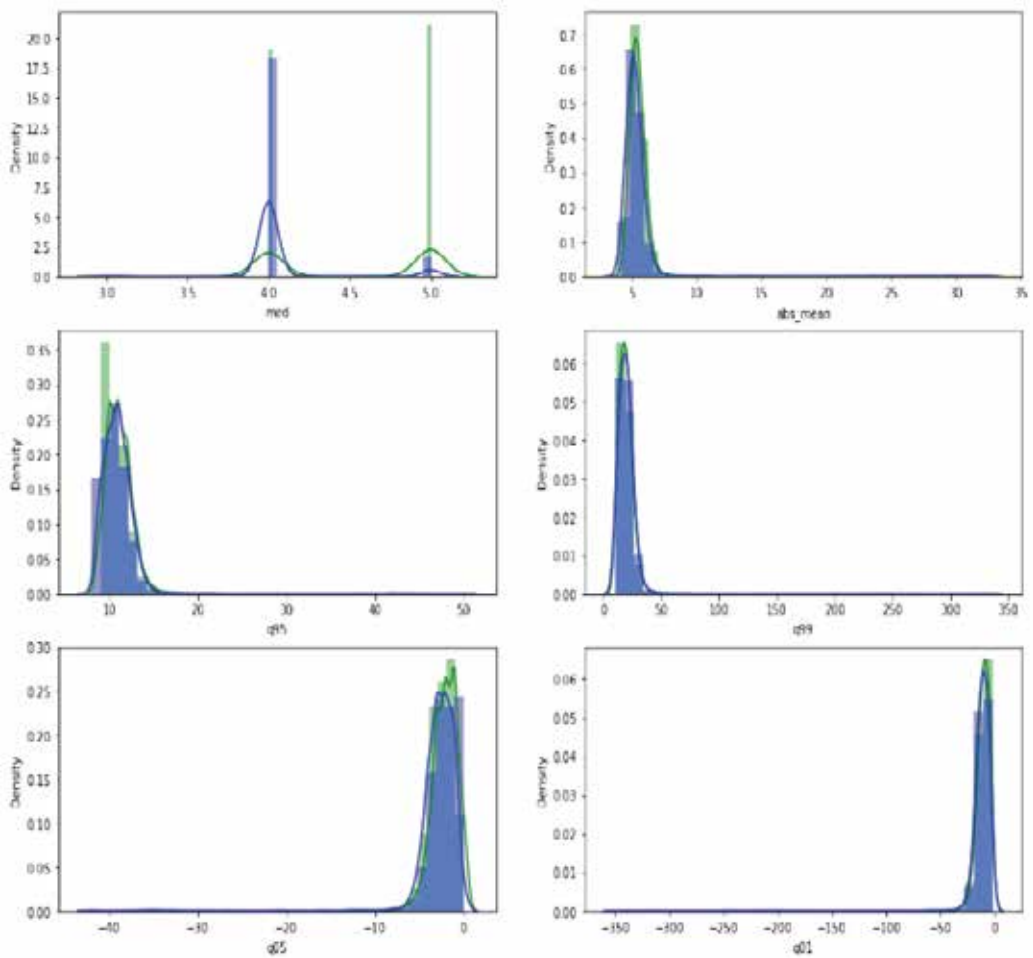


Figure 8

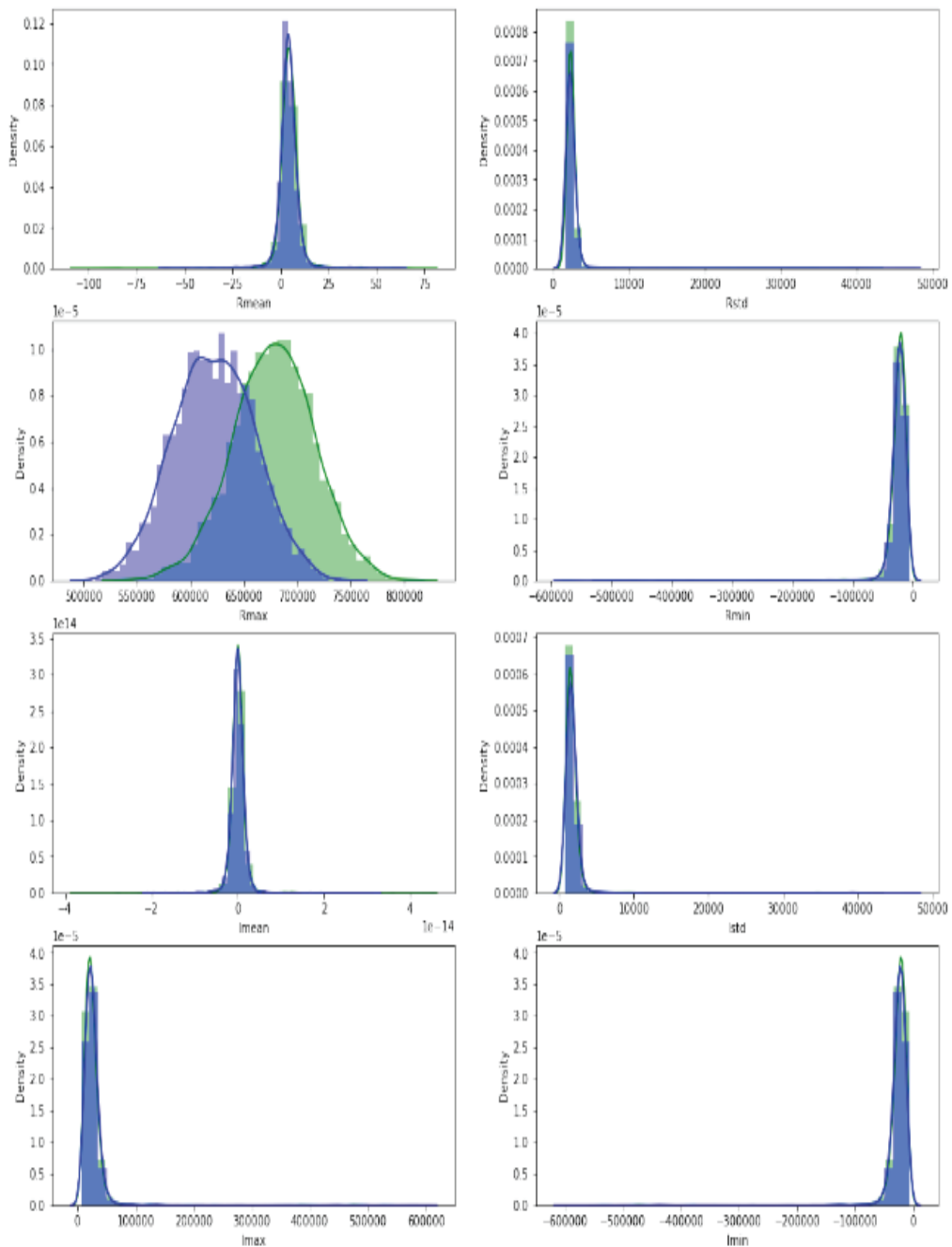


Figure 9

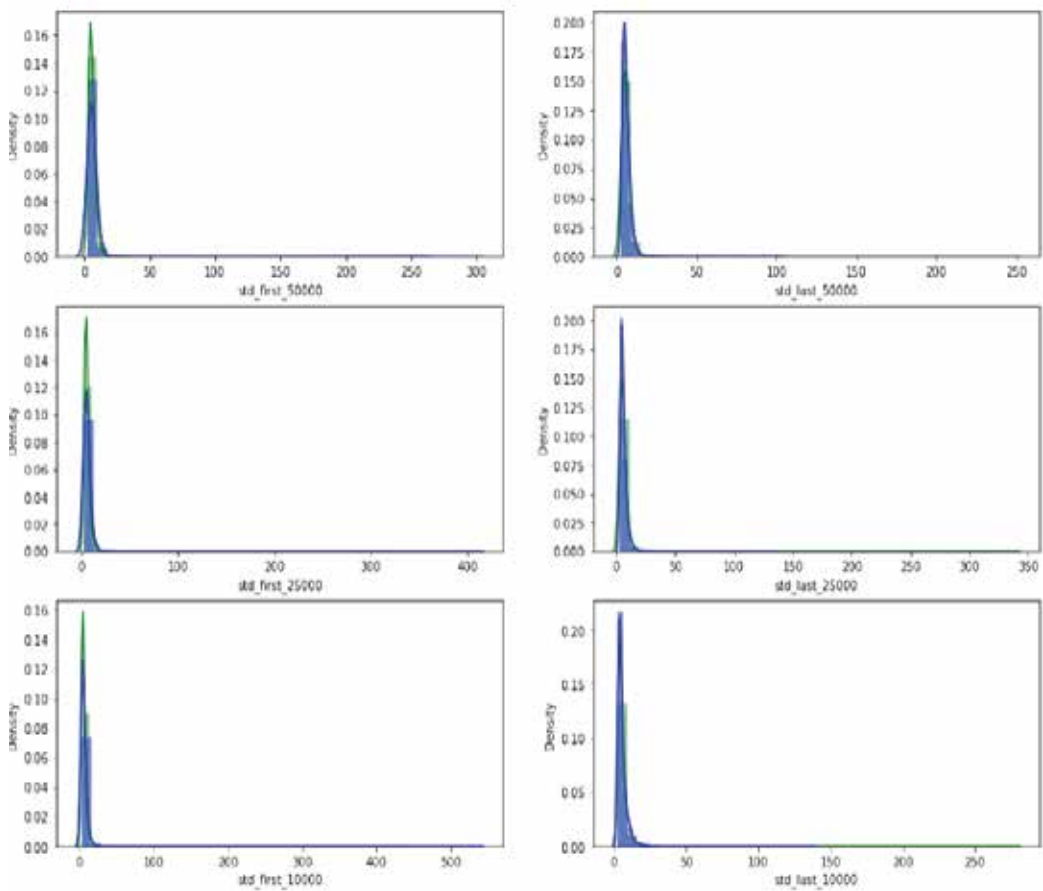


Figure 10

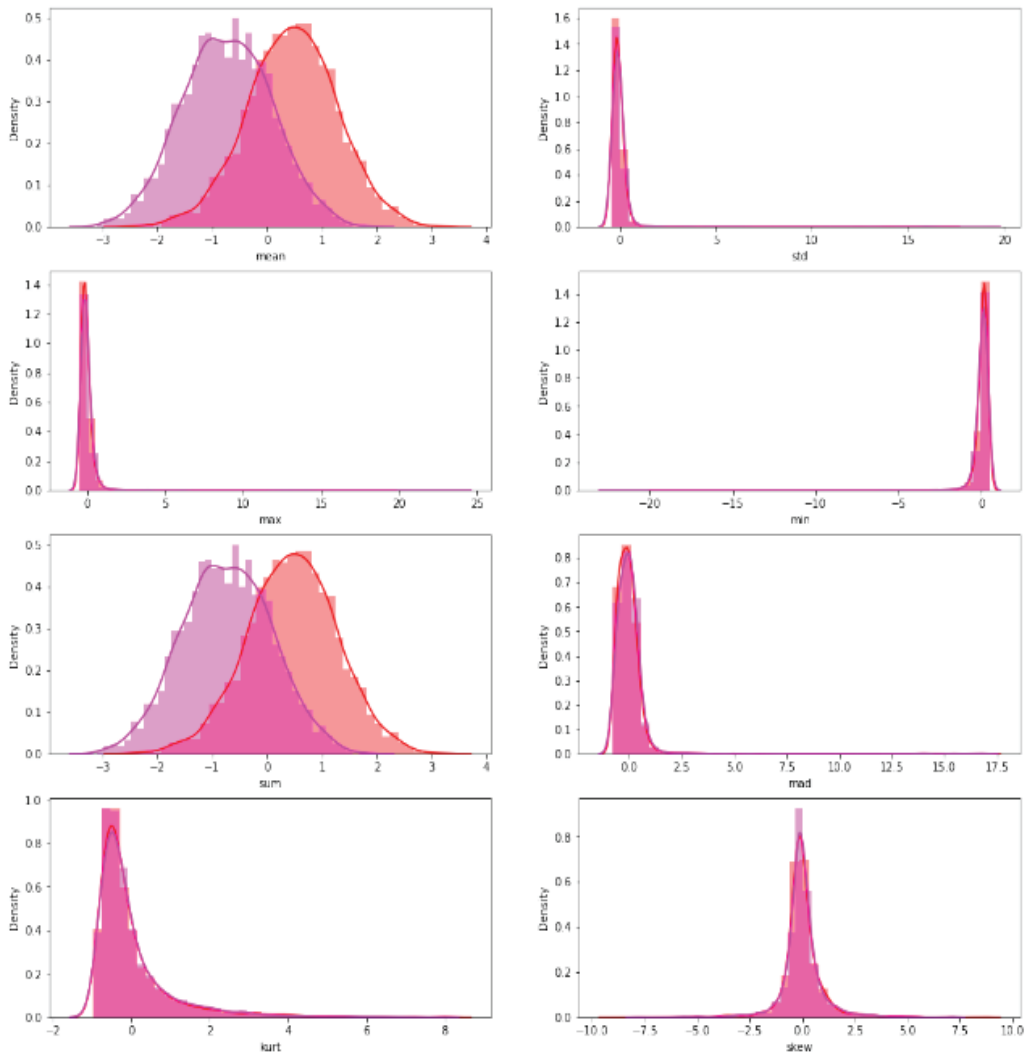


Figure 11

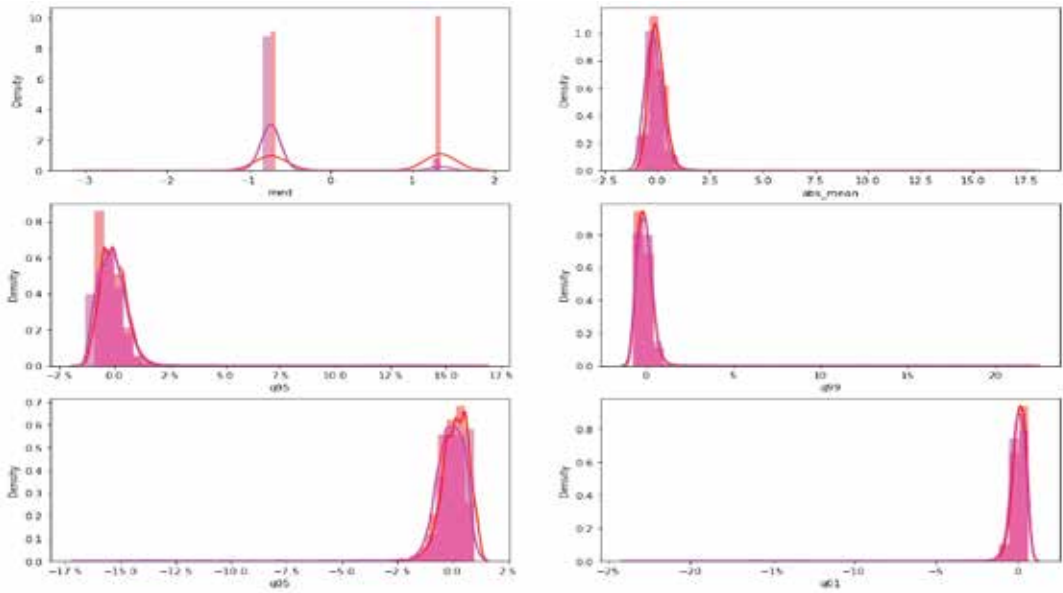


Figure 12

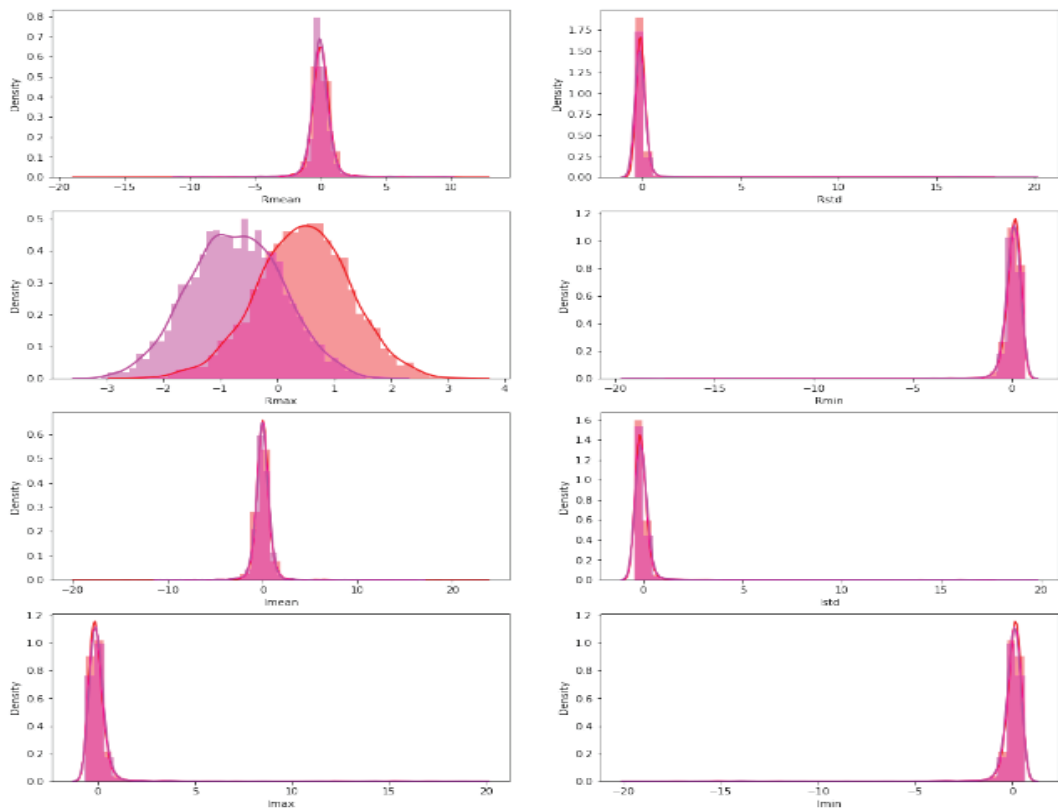


Figure 13

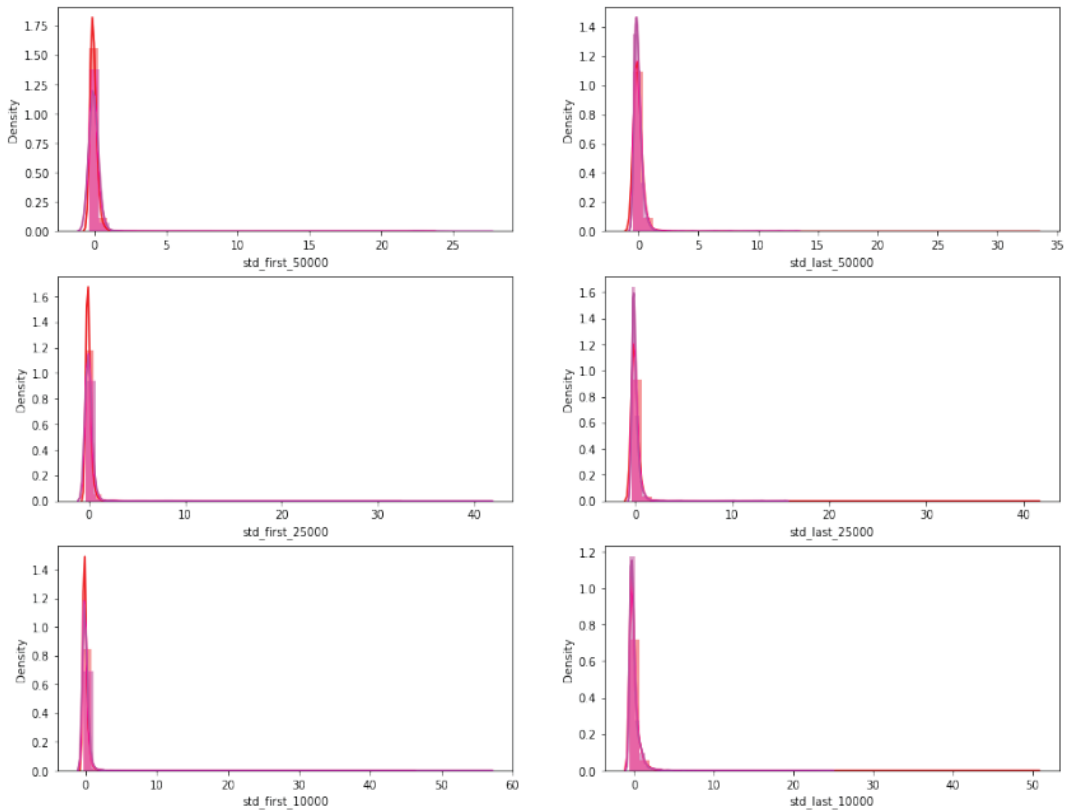
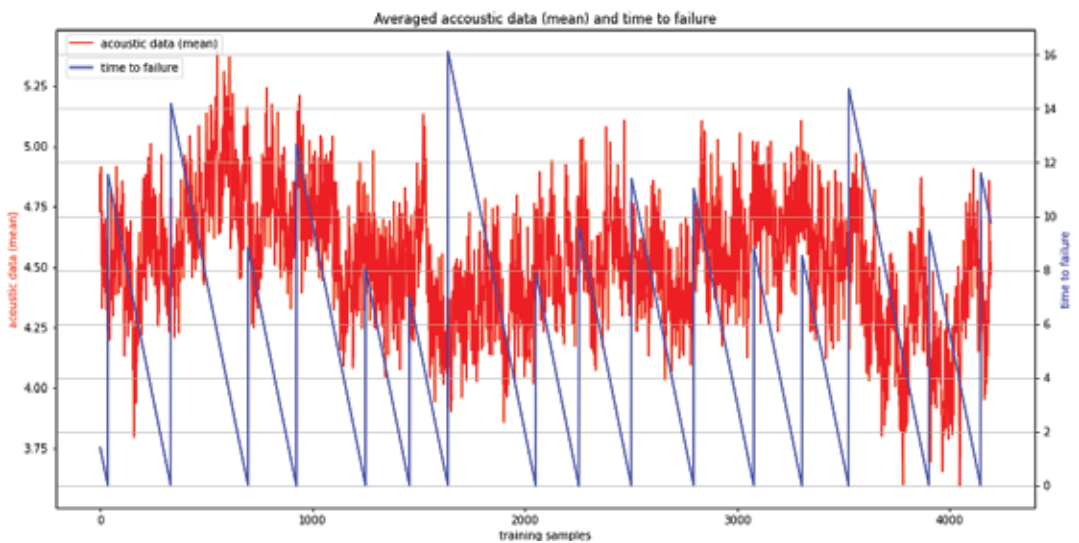


Figure 14



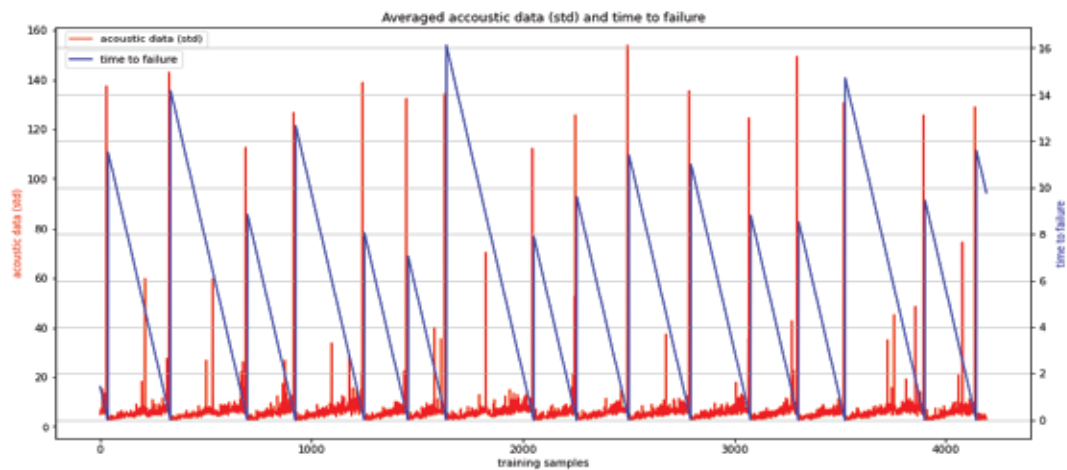


Figure 16

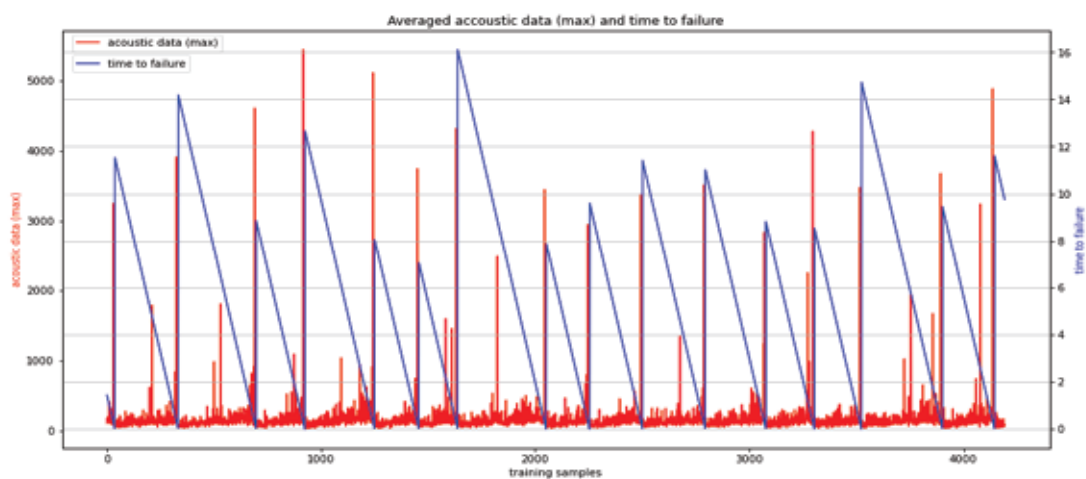


Figure 17

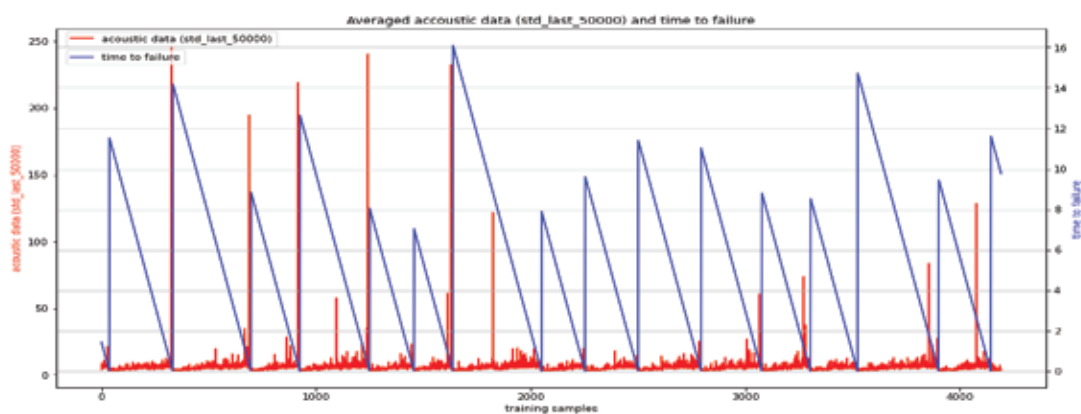


Figure 18

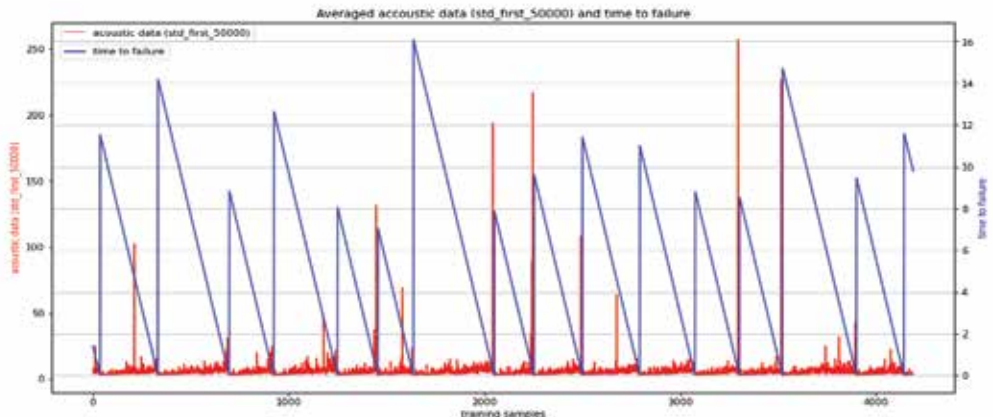


Figure 19

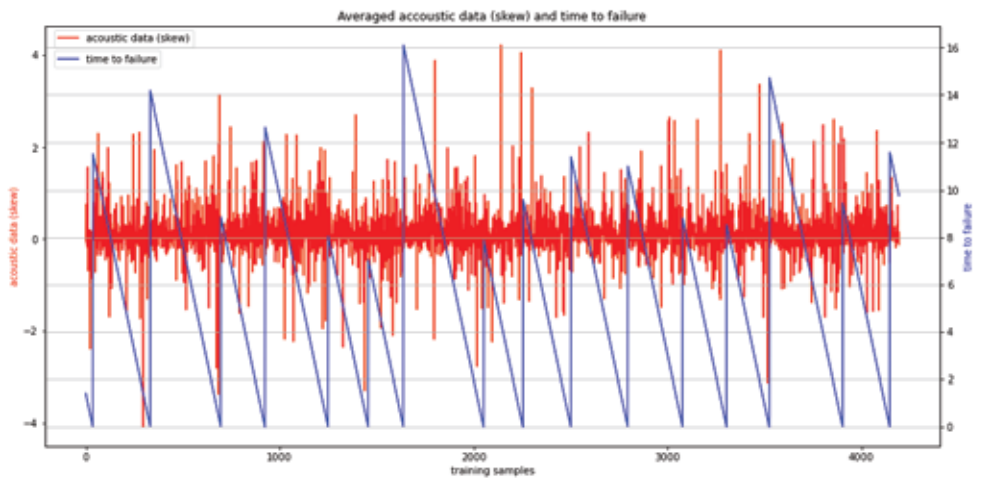


Figure 20

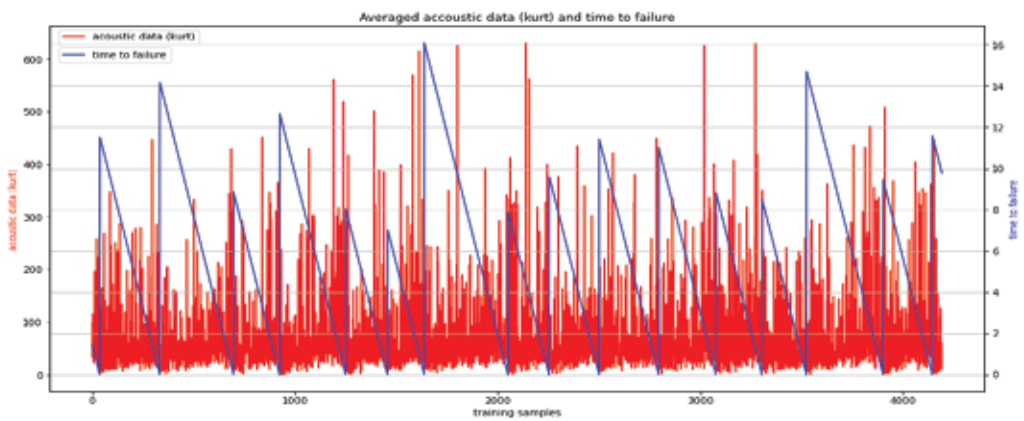


Figure 21

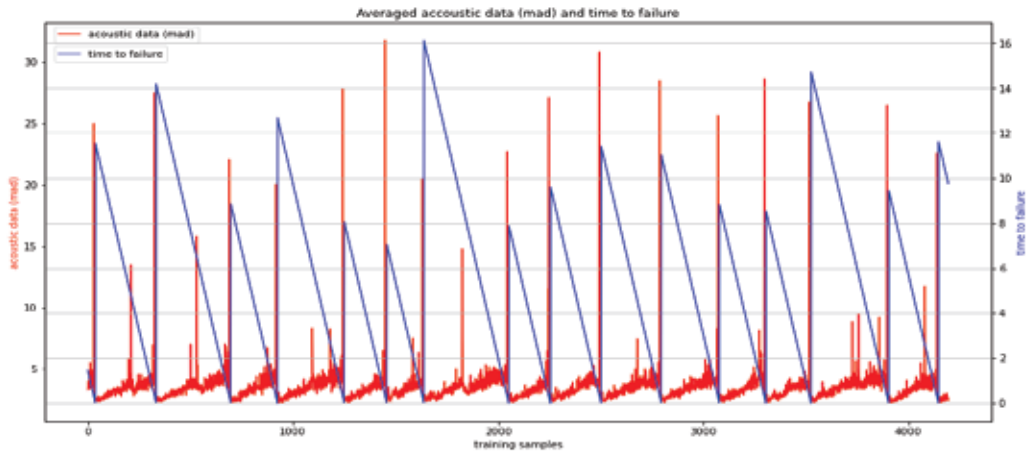


Figure 22

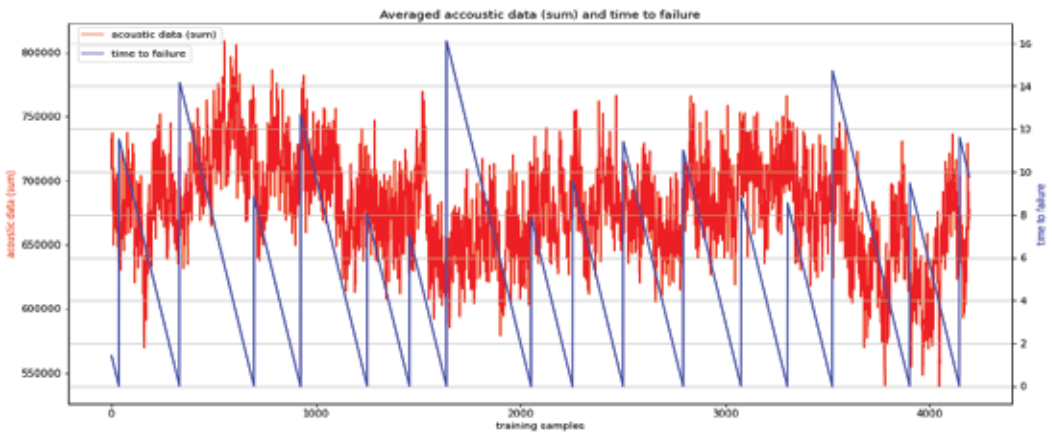


Figure 23

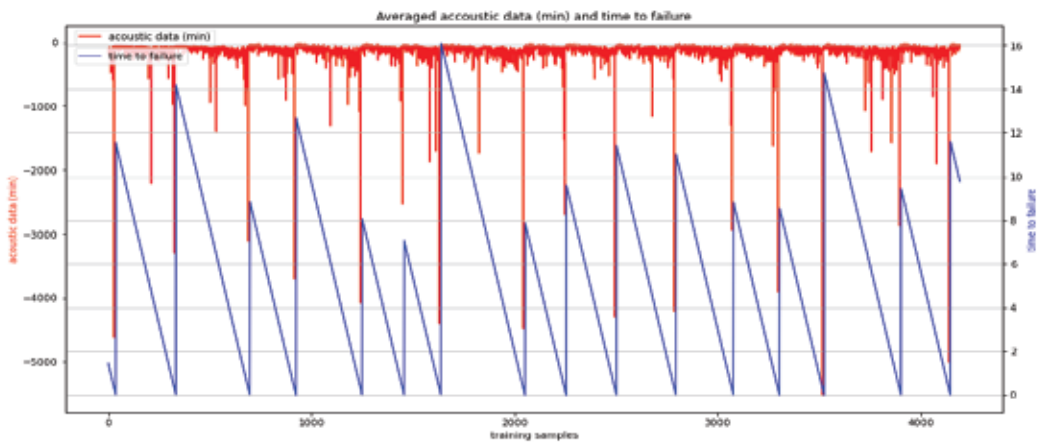


Figure 24

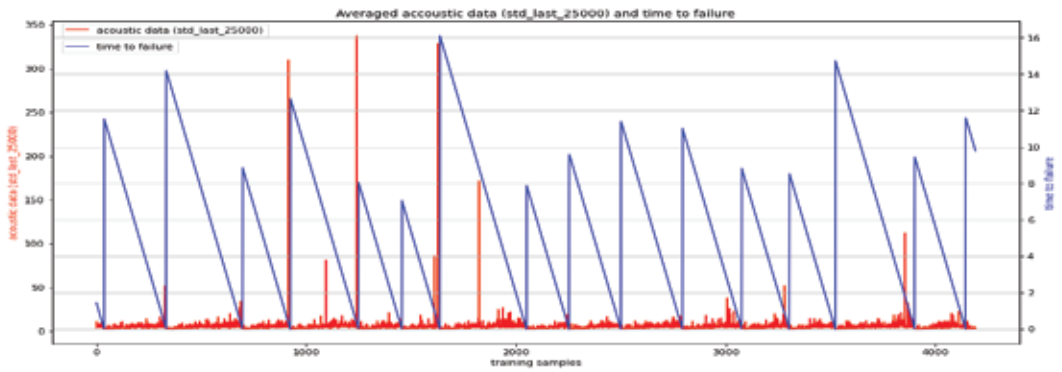


Figure 25

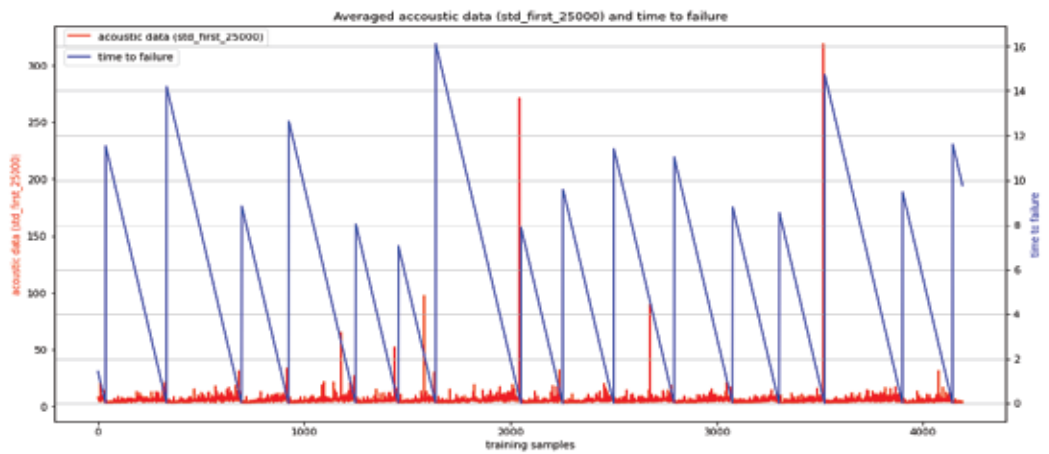


Figure 26

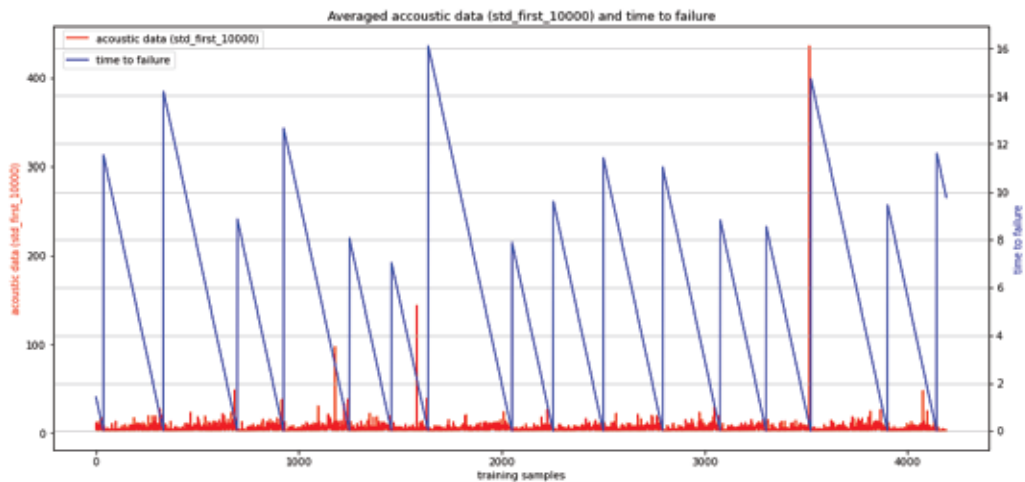


Figure 27

We broke down the conveyance of the accumulated highlights and furthermore the opportunity to disappointment and the totaled highlights on a similar chart

Referenc

1. Afzal, S. a. (2019). Load balancing in cloud computing – A hierarchical taxonomical classification. *Journal of Cloud Computing* , 1-24.
2. Al Nuaimi, K. M.-J. (2012). A survey of load balancing in cloud computing: Challenges and algorithms. *second symposium on network cloud computing and applications. IEEE*, 137-142.
3. Ala'Anzy, M. a. (2019). "Load balancing and server consolidation in cloud computing environments: A meta-study. *IEEE Access*.
4. Arulkumar, V. a. (2021). Performance analysis of nature inspired load balancing algorithm in cloud environment.". *Journal of Ambient Intelligence and Humanized Computing*, 3735-3742.
5. Asghari, A. M. (2021). "Task scheduling, resource provisioning, and load balancing on scientific workflows using parallel SARSA reinforcement learning agents and genetic algorithm.". *The Journal of Supercomputing*, 77.
6. Chen, Y.-C. a.-L. (2018). "The novel cloud application technology with virtual platform.". *MATEC Web of Conferences*. (p. 01007.). EDP Sciences.
7. Geetha, P. a. (2017, August)). "A comparative-study of load-cloud balancing algorithms in cloud environments. *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 806-810). IEEE.
8. Ghahramani, M. H. (2017). Toward cloud computing QoS architecture: Analysis of cloud systems and cloud services. *IEEE/CAA Journal of Automatica Sinica*, 6-18.
9. Gundu, S. P. (2020). A. Real-Time Cloud-Based Load Balance Algorithms and an Analysis. *SN COMPUT. SCI*.
10. Ibrahim, I. M. (2021). Task scheduling algorithms in cloud computing: A review.". *Turkish Journal of Computer and Mathematics Education* , 1041-1053.
11. Jena, U. K. (2020). Hybridization of meta-heuristic algorithm for load balancing in cloud computing environment. *Journal of King Saud University-Computer and Information Sciences*.
12. Khan, M. A. (2020). A Machine Learning Approach for Blockchain-Based Smart Home Networks Security. *IEEE Network* .
13. Kumar, M. &. (2019). PSO-based novel resource scheduling technique to improve QoS parameters in cloud computing. *Neural Computing and Applications*, 1-24.
14. Lyu, F. R. (2020). SoSA: Socializing Static APs for Edge Resource Pooling in Large-Scale WiFi System. *In IEEE INFO-*

- COM 2020-IEEE Conference on Computer Communication*, 1181-1190.
15. Mallikarjuna, B. a. (2019). The role of load balancing algorithms in next generation of cloud computing. *J. Adv. Res. Dyn. Control Syst.*, 1715-1733.
 16. Mortada, W. I. (2020). Recent developments and applications of cloud point extraction: A critical review. *Microchemical Journal*.
 17. nzinger, C. N. (2014). MADCAT: A methodology for architecture and deployment of cloud application topologies. in *2014 IEEE 8th international symposium on service oriented system engineering*, 13-22.
 18. Panda, S. K. (2019). Load balanced task scheduling for cloud computing: A probabilistic approach. *Knowledge and Information Systems*, 1607-1631.
 19. Panda, S. K. (2019). *Load balanced task scheduling for cloud computing: A probabilistic approach*. Knowledge and Information Systems.
 20. Sambit Kumar Mishra a, . B. (2020). Load balancing in cloud computing: A big picture. *Journal of King Saud University* —.
 21. Wagle, S. S. (2015). "Cloud service providers ranking based on service delivery and consumer experience.". *IEEE 4th International Conference on Cloud Networking (CloudNet)*, 209-212.
 22. Yoo, C. S. (2011). Cloud computing: Architectural and policy implications. *Review of Industrial Organization*, 405-421.