DECENTRALIZED ARTIFICIAL INTELLIGENCE USING BLOCKCHAIN

Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering/Information Technology

By

Arpit Gurudutt Mishra 181370
Purav Vashisht 181287

Under the supervision of

Prof. Dr. Vivek Kumar Sehgal

to

ACKNOWLEDGEMENT

Firstly, We would like to express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the project work successfully

We are really grateful and wish our profound indebtedness to Dr. Vivek Sehgal, Professor & HOD of the Department of CSE & IT, Jaypee University of Information Technology, Waknaghat.

Deep Knowledge & keen interest of my supervisor in the field of “Deep Learning and AI” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would also generously welcome each one of those individuals who have helped us straightforwardly or in a roundabout way in making this project a win.

In this unique situation, I might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated my undertaking.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

Regards,

Arpit Gurudutt Mishra 181370
Purav Vashish 181287
CSE | JUIT

Table of content :

Department of Computer Science & Engineering and Information Technology
Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh

Certificate

Candidate's Declaration

I hereby declare that the work presented in this report entitled **" Decentralized Artificial Intelligence Using Blockchain"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from August 2021 to December 2021 under the supervision of **(Supervisor name)** (Designation and Department name).
The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature)
Student Name, Rollno.
Arpit Gurudutt Mishra 181370
Purav Vashisht 181287

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature)
Supervisor Name: Prof. Dr. Vivek Kumar Sehgal
Designation: Professor and Head of Department
Department name: Computer Science and Engineering
Dated: 06-12-2021

List of Abbreviations:

AI - Artificial intelligence

ML - Machine Learning

DAI - Decentralized Artificial Intelligence

GPVS - General Purpose Vision System

GPU - Graphical processing unit

List of Figures :

Chapter-1 Introduction

Bitcoin creation has delivered every other increase withinside the Internet world. In a totally brief time, many Blockchain structures have come into being successively, whose decentration, consensus mechanisms, smart contract, and different traits cause them to be relevant to diverse fields inclusive of finance, education, medical, era, etc. The consensus mechanism in the blockchain is an important part of Blockchain technology. A perfect consensus mechanism is a vital part in the operation of the Blockchain System.

The never ending development of consensus mechanism inclusive of PoW (proof-of-work), PoS (proof-of-stake), DPoS (delegated-proof-of-stake), and PBFT (Practical Byzantine Fault Tolerance) has brought about the evolution of Blockchain era to Blockchain 3.0.

Blockchain is considered as an effective solution to security, but even blockchain can be prone to attacks and threats, even blockchain based decentralized platform have experienced significant security threats.

Blockchain has consistently tracked down places in conversations about transactions as for digital forms of money. In any case, the non-monetary use instances of blockchain in computerized character the board and store network the executives have featured more with regards to its latent capacity. Thus, the extension for consolidating blockchain and AI arose in this manner as various use instances of blockchain began acquiring consideration. Prior to observing how they are connected to one another, let us comprehend the terms.

Blockchain is a disseminated advanced record shared across peers in the organization. The companions or hubs settle on exchanges that you need to add to the blockchain network. The exchanges on the organization are put away on blocks that have exceptional hash esteems close by time stamps for checking honesty. The association of squares to one another as a chain gives the justification for the term 'blockchain'. The chain of associated distributed organizations is essentially unchanging and offers the ideal security from information adjustment.

Computerized reasoning is the capacity for the recreation of human knowledge using machines. The blockchain computerized reasoning condition to a great extent relies upon

the capacities of AI for empowering innovative arrangements with intellectual characteristics. The essential objective of computerized reasoning spotlights on decreasing human mistakes while guaranteeing quicker tasks. Hence, you can plainly see how both AI and blockchain expect to make processes quicker. Utilizing the two of them together certainly presents some intriguing possibilities for growing the utilizations of blockchain across different areas.

Machine Learning has as of late empowered enormous advances in artificial intelligence however these outcomes can be exceptionally unified. The huge datasets required are for the most part restrictive; forecasts are frequently sold on a for each inquiry premise; and distributed models can immediately become out of date without work to procure more information and keep up with them To address centralization in AI, systems to share machine learning models on a public blockchain while keeping the models allowed to use for deduction have been proposed. One model is Decentralized and Collaborative. For instance, Machine learning on Blockchain from Microsoft Research. The work centers around the depiction of a few potential impetus systems to urge members to add information to prepare a model.

The mix of AI and blockchain could present revolutionary developments later on. An itemized comprehension of existing patterns related with the blockchain and AI relationship could help in recognizing the explanations behind thinking about the combo. As indicated by the International Data Corporation, worldwide costs on AI expanded to nearly $57.6 billion by 2020. Curiously, around 51% of organizations utilized the AI and blockchain blend. Existing patterns of applying AI in the blockchain would improve blockchain by presenting the accompanying functionalities,

- Information insurance
- Shrewd processing power
- Information adaptation
- Formation of assorted informational indexes

Blockchain could likewise give a superior stage to getting AI. It can help in following the dynamic cycles in AI. Blockchain could help in exact documentation of every

information and variable associated with the decision-production of AI calculations. Besides, blockchain AI intermingling is likewise positive on the grounds of guaranteed upgrades in blockchain productivity. For instance, Artificial Intelligence could help in computerizing different parts of blockchain the executives, for example, review trail observing.

By utilizing progresses in AI, forecast markets, and blockchain stages, we can show the capacities of another structure to gather tremendous measures of information, permit supporters of conceivably benefit, and host a common machine learning model as a public asset. The model can be cooperatively prepared by numerous benefactors yet stay open furthermore, free for others to utilize the model for surmising. This is achieved with a few configurable parts:

- The incentive mechanism
- The data handler
- The AI model

A brilliant agreement is made and introduced with decisions for these parts. It then, at that point, acknowledges "add information" activities from members, with the motivating force component potentially setting off transactions or permitting different activities. Adding information includes approval from the motivating force component, putting away in the information overseer, lastly calling the update strategy on the model's contract, calling the anticipate work given to accommodation in the model's smart contract code.

1.1 Blockchain Technology

The design of blockchain basically involves the information layer, network layer, agreement layer, motivating force layer, contract layer, and application layer

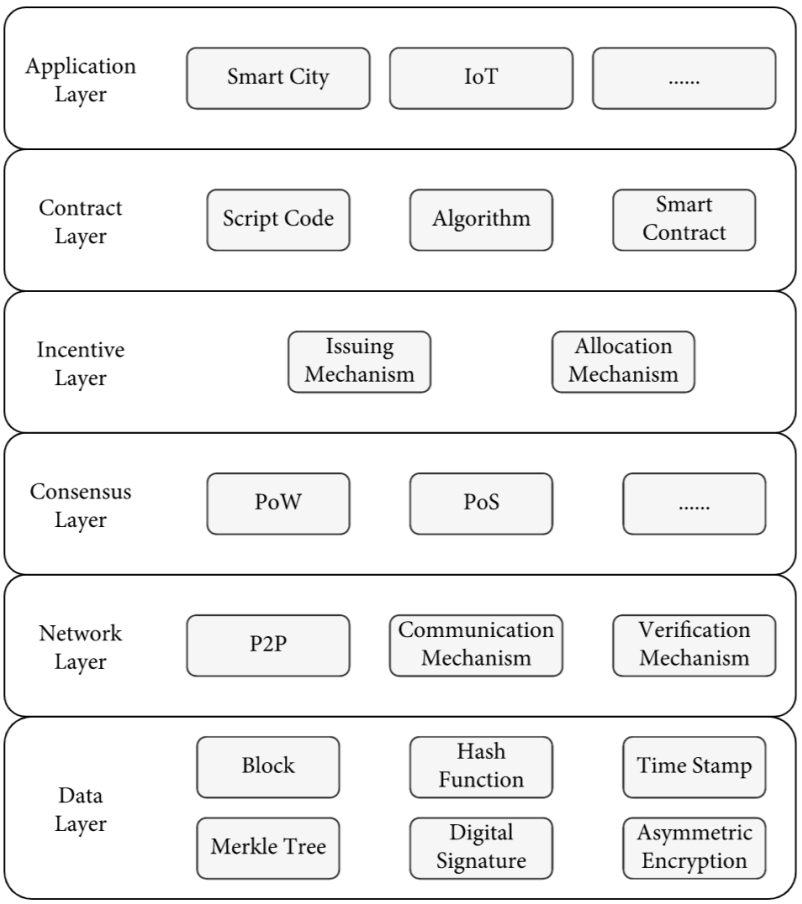


Fig. 1

Data layer chiefly centers around the information structure, including the hash work, computerized signature, Merkle tree, non-symmetric encryption, and different advancements. The main design of the information layer is the block, and the block construction is displayed in Figure 2. A block comprises both the body and the head. The blockchain contains the Merkle root, timestamp, and hash worth of the current square and past block. The blockchain mostly incorporates exchange data and the Merkle tree. Every exchange is endorsed by the exchange's initiator and afterward handled and checked by the excavator. The checked exchange is implanted in the square. The hash worth of each exchange is joined two by two to ascertain the hash, and afterward, the subsequent hash esteem is consolidated two by two to work out the hash esteem again until the Merkle root, which is recorded in the blockchain. Each change to the data about each exchange put away on the blockchain influences the Merkle root. Along these lines, the sealing of

blockchain can be figured out. Each block moreover stores the hash worth of the past block and timestamp, bringing about a period arranged chain.

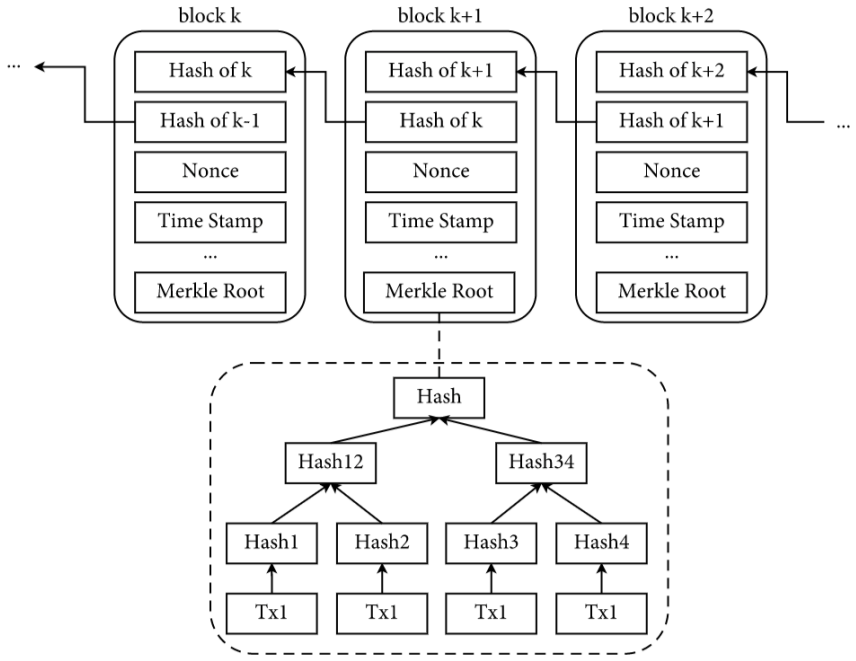


Fig. 2

Network layer basically contains P2P organization, plan of the information correspondence instrument, and information confirmation system. There is no brought together server in the blockchain. All messages spread between hubs in a shared way. All hubs keep up with the blockchain together. One hub creates another square and sends it to different hubs. Different hubs store the duplicate of the square after confirmation. Ensuing squares will likewise be created based on this square. Thus, everything hubs can keep a base record.

The agreement layer basically incorporates different agreement calculations. The agreement calculation is utilized to figure out which hub can add new squares to the principle chain. Normal agreement calculations incorporate PoW, PoS, and PBFT.

Incentive layer for the most part incorporates some motivator measures. There is no concentrated server in the blockchain, so the protected activity of the blockchain relies upon the dynamic cooperation of every hub. As of now, the regularly utilized impetus

measures incorporate (1) the related compensation for each square's accounting right and (2) the help charge for every exchange. With the advancement of blockchain, the plan of the impetus layer of blockchain in what's to come isn't simply restricted to financial rewards yet additionally to accomplish shared objectives.

Whereas the contract epitomizes a few contents, calculations, and shrewd agreements to help the programmable elements of blockchain. Through the preset standards and conditions, it tends to be naturally executed without an outsider. This is the establishment of blockchain trust. The last part is the application layer. It contains a wide range of blockchain applications, including finance, law, review, and medical services.

1.2 Smart Contract

Szabo first defined the Smart Contract in 1994 smart contract have not been taken on for an enormous scope in view of the absence of a dependable execution climate before the presentation of blockchain innovation. With the fast improvement of disseminated record innovation, particularly the huge scope arrangement and utilization of blockchain, individuals focus closer on savvy contract, which is the key and significant piece of dispersed record innovation.

Smart Contract is a sort of a convention that would self be able to execute, self authorize, self confirm, and self limit the execution of its directions. It permits exchanges to be executed between untrusted or mysterious gatherings without the requirement for a confided in outsider. These exchanges are discernible and irreversible. Savvy contract comprises worth, address, capacity, and state. The exchange is taken as info, the relating code is executed, and a result occasion is set off; then, at that point, the state changes as per the useful rationale . All gatherings settle on the subtleties of the brilliant agreement ahead of time, including situations that trigger agreement execution, state change rules, and obligation regarding break of agreement. Then, at that point, the shrewd agreement is conveyed on the blockchain as a code. From that point forward, when the prerequisites are fulfilled, the savvy agreement will be set off and naturally executed.

Ethereum is the most well known stage for the improvement of shrewd agreement. The code of Ethereum brilliant agreement is written in stack-put together bytecode language and runs with respect to Ethereum virtual machine (EVM). Robustness and Serpent are regularly used to compose brilliant agreements. Hyperledger texture can likewise convey smart contract, which is classified "chain code." It is the single communication channel with the blockchain and the main hotspot for exchange age. Chain code is normally evolved utilizing Go or Java.

1.3 Consensus Mechanism

There is no trust connection between every node in the blockchain. It is important to facilitate every autonomous node to share data in such network. Subsequently, the network framework will conclude who is the following clerk through pertinent conventions to arrive at an agreement of which is the agreement component. The quintessence of the agreement system is to tackle the issue of decentralized trust. It is a fundamental innovation for the free activity of the blockchain. In the consistent working of a blockchain framework, a decent agreement system has an extremely huge impact. The blockchain can effectively arrange and develop a predictable blockchain structure utilizing a successful agreement component.

There are two kinds of agreement calculations. One is for non-Byzantine shortcoming, like RAFT and Paxos. The other is for the Byzantine general issue, like PoW, PoS, DPoS, and PBFT. There are two methods for managing Byzantine issue. One is to restrict the likelihood of malignant conduct by expanding the expense of doing insidious, like PoW and PoS. The PoW calculation will consider processing power cost, and the PoS calculation will consider stakes cost. Another way is to plan specific guidelines. Regardless of whether there are sure noxious hubs, any remaining hubs can in any case arrive at an agreement, for example, the pragmatic Byzantine shortcoming open minded calculation. A few normal agreement calculations are portrayed as follows:

(1) Proof of Work: the PoW calculation was first proposed in Bitcoin, and its center thought is the opposition of hub processing power. The excavator can have the accounting right by burning-through a great deal of processing ability to ascertain a hash esteem that meets the necessities. The square will be added to blockchain after different hubs approving it. Then, at that point, the accounting hub will be compensated. In the PoW agreement instrument, it takes a great deal of assets for pernicious hubs to annihilate the framework (control in excess of 50% of hubs). In this manner, it can restrict the malevolent conduct of malignant hubs. PoW can be decentralized, and hubs can enter and leave openly. However, clearly, it will cause a misuse of assets and low effectiveness.

(2) Proof of Stake: the PoS calculation is a choice to settle the misuse of assets in the PoW calculation. It lessens the trouble of mining because of the number and season of tokens taken by every hub. Partially, it abbreviates an opportunity to arrive at agreement and evades a great deal of misuse of assets in the PoW calculation. However, simultaneously, PoS benefits rich excavators and may prompt close to syndication. In this manner, blockchain projects utilizing the PoS calculation normally need to run the PoW agreement calculation for a while and afterward convert to PoS to forestall countless stakes collecting in few hubs.

(3) Delegated Proof of Stake: the DPoS agreement calculation enhances the premise of the PoS calculation. The agreement interaction no longer requires all taking part hubs to battle for the accounting privileges, yet pick a few agents through casting a ballot. It extraordinarily works on the effectiveness of agreement.

1.4 Advantages of combining AI with Blockchain

Obviously two distinct advances have special qualities when working freely. Nonetheless, joining the best of both for the bigger great is totally obvious in a blockchain and AI combo. Here are a portion of the striking benefits of joining AI with blockchain.

1.5 Encryption Advantages

AI and blockchain could give a considerable lift to enhancements in encryption. As a matter of first importance, AI has impressive potential concerning security. Late advancements in AI have been centered around creating calculations equipped for working with information in an encoded state. This is clearly a security hazard while blockchain security calculations can make a strong mediation with data put away in encoded structure.

Blockchain AI applications could offer the advantage of putting away exceptionally touchy individual information. With the ideal and savvy handling draws near, the information could help in opening comfort and worth. For instance, savvy medical care frameworks could guarantee exact medical services schedules by filtering clinical records with guaranteed security.

1.6 Simplicity of Blockchain Management

The advantage of better administration is a conspicuous justification behind which think about AI and blockchain blends. We have consistently had quick PCs, in spite of the fact that with next to no sign what to do until furnished with explicit directions. In this manner, using blockchain on computers could suggest the requirement for a lot of handling power. Hashing calculations for mining blocks follow an animal power approach by attempting various mixes of characters for tracking down the appropriate option for checking a particular exchange. Along these lines, it is unmistakably apparent that managing blockchain most certainly takes a ton of handling ability to do each interaction. Presently, the blockchain and AI mix could fill in as a solid answer for resolving this issue. Simulated intelligence can help in making a shift away from the animal power approach, accordingly overseeing assignments with shrewd and sensible methodologies. Artificial intelligence calculations are custom-made for learning with each collaboration inside the space of seconds that a human master would require numerous years to learn. Blockchain is famous for its decentralization and straightforwardness. Subsequently, it gives the ideal instrument for stripping the layers of

intricate AI calculations to comprehend their dynamic cycles. The choices by AIs could be hard for people to comprehend. Be that as it may, there are circumstances where AI-based choices should be brought to review, principally for confirming exactness. The most conspicuous illustration of Walmart shows that it inputs very nearly one month of value-based information all through the stores in its AI frameworks, settling on choices in regards to loading or area of items.

The blockchain and AI mix can work completely for this situation with the benefits of the datapoint-to-datapoint approach for recording choices. Accordingly, it additionally presents productive freedoms for working on the believability of AI. The developing trust of individuals in AI could prompt new walks in straightforwardness and experiences into the personalities of robots.

1.7 Machine Learning Models

In this part, we layout a few models decisions of AI model
for use with Decentralized and Collaborative AI on Blockchain as proposed. The model engineering picked relates near the motivation system
picked. In this work, we will investigate models for the Self-Assessment impetus system as it requests to the decentralized idea of public blockchains in that
an incorporated association will not have to keep up with the IM, for instance, by financing it.

For our examinations, we mostly consider managed classifiers since they
can be utilized for some applications and can be effortlessly assessed utilizing test sets. In request to keep exchange costs low, we initially propose to use the work in the
Steady Learning space by utilizing models able to do productively refreshing
with one example. Exchange expenses, or "gas", are
significant for most open blockchains as a method for paying for the calculation cost for executing a savvy contract.

Naive Bayes

The model initially is a Naive Bayes classifier for its relevance to many kinds
of issues. The Naive Bayes classifier accepts each element in the model

is free, this is the thing that helps makes calculation quick when refreshing and foreseeing. To refresh the model, we without a doubt need to refresh a few considers such the quantity of information focuses seen, the occasions each element was seen, the number of times each element was seen for each class, and so on While anticipating, all of
these counts are utilized for the highlights introduced in the scanty example to process the most probable class for the example utilizing Bayes' Rule [8].

Nearest Centroid

A Nearest Centroid Classifier registers the normal point (or centroid) of all focuses in a class and orders new focuses by the mark of the centroid that they are nearest to. They can likewise be effortlessly adjusted to help different arrangements (which we don't accomplish for this work). For this model, we follow along of the centroid for each class and update it utilizing the total moving normal technique. Along these lines we additionally need to record the quantity of tests that have been given for each class. Refreshing the model with one example needs to refresh the centroid for the given class yet not for different classes. This model can be utilized with thick information portrayals.

Perceptron

A solitary layer perceptron model is valuable straight model for parallel grouping. We assess this model since it very well may be utilized for scanty information like text as well as thick information. The Perceptron's update calculation just updates the loads assuming that the model as of now characterizes the example as inaccurate. This is useful for our framework since it should help stay away from overfitting. The model can be proficiently refreshed
simply by adding or taking away, contingent upon the example's mark, the qualities for the elements of the example with the model's loads.

Datasets

Three datasets were picked as instances of issues that would profit from community oriented situations where numerous supporters can work on a model all together
to make a common public asset. In every situation, the clients of an application

that would utilize such a model advantage by straightforwardly affecting working on the model they habitually use and not depending on a brought together power to have and control the model.

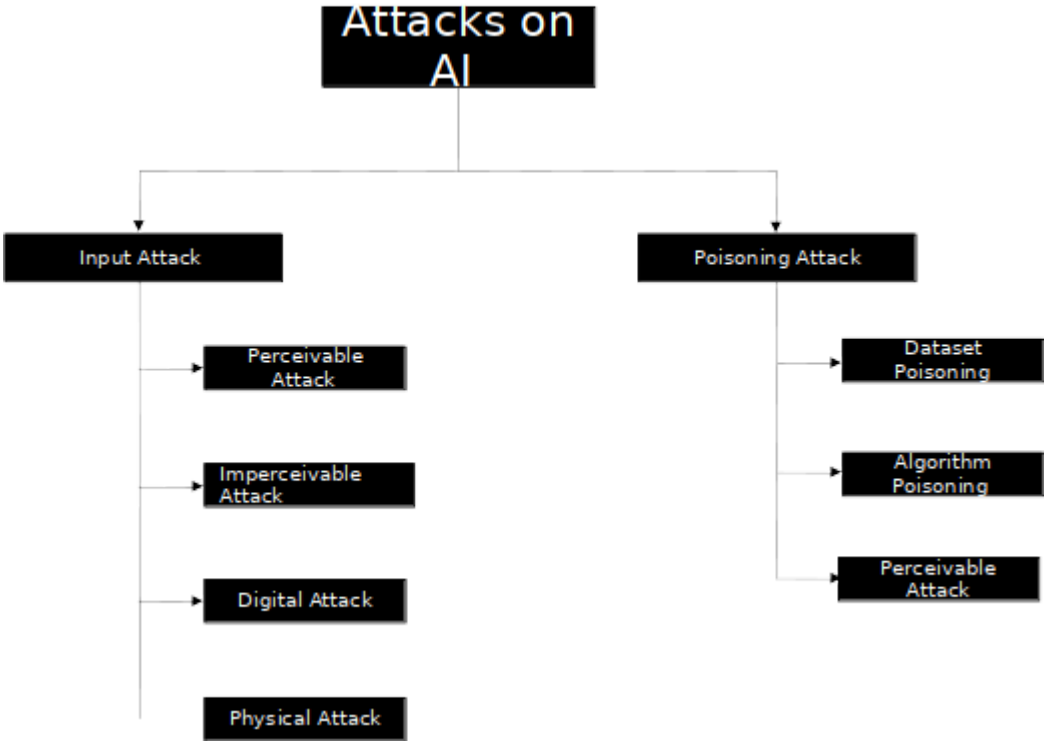


Fig. 3: Attacks on AI

AI has changed the way we live. Automate is going on at each level and area at an undeniably quicker pace. Simulated intelligence is progressing significantly,

Furthermore, it has changed everything socially, monetarily, and strategically. Artificial intelligence has significantly progressed the application regions like medical services, business, training, independent vehicles, the movement business, web-based media, and horticulture. Expanded reception of AI for basic undertakings makes it more defenseless against assaults, as specific application regions, for example, in medical care, military and common society are becoming appealing objective regions for the assault.

Figure 1, portrays the various classes of assaults on AI frameworks. Input assaults are those wherein contribution to the AI System is controlled to change the result, as for the situation of an irritated picture. It doesn't have to have an adulterated AI framework. Input assaults fall under four classes: discernible, imperceivable, advanced and actual assaults. Distinguishable assaults are those on actual substances and are noticeable to the natural

eye. Assaults on physical or on the other hand computerized elements that are undetectable to human faculties are known as intangible assaults.

Advanced assaults are directed on computerized information like pictures, recordings, archives, and documents also are for the most part imperceivable. On account of actual assaults, the objective is actual items. In the vast majority of the situations, actual assaults are simple and distinguishable. In a harming assault, the aggressor means to hurt the AI model so when it is utilized, it has normal shortcomings that empower them to control it without any problem.

The learning calculations in dataset harming learn a model by distinguishing designs in the harmed information, bringing about a disturbed learning process. Controlling the calculation by distinguishing the shortcoming in it is known as calculation harming. Unified learning has the danger of such assaults as the client controls information and calculations. For this situation, despite the fact that a model has been entirely prepared with a totally qualified data set and not really set in stone to be non-harmful, it can in any case be refreshed with a poisonous model at various focuses along the advancement interaction. These actions would be those that identify and redress expected wellsprings of information, calculation, model, and info harming that outcome in 'degenerate' yields.

The increment in ubiquity and need for computerized reasoning applications, it is being examined by basic applications wherein blind confidence in a savvy application can put human lives in danger. This review centers around the AI application regions for example, the recommender framework, IoV based shrewd city sending, energy market, AIbased medical services applications and portable edge organization, in which Blockchain innovation is adding trust and straightforwardness. Beginning around 2015, the examination in consolidating Blockchain with AI has started. Concerning development examples, intensity, and the social, scholarly, and hypothetical setting of the field, this review is expected to give bits of knowledge into the headway of what's more arising patterns in coordinating two compelling advances, Blockchain and AI. To begin with, the examination analyzes patterns in distributions and reference information from 2015 to April 2021 to lay out the advancement in the recorded area (i.e., research patterns). The concentrate then, at that point, distinguishes the critical diaries and spaces

of examination that are generally pertinent to the field's development and the driving creators and areas that add to investigate on incorporating Blockchain and AI(i.e., research virtu).

Finally, the review remarks on the potential exploration relationship and scholarly and scientific literature for the research of Blockchain and AI integration since 2015. The insights of this study are as follows:

1. To conduct a Bibliometric study for the integration of Blockchain and artificial intelligence considering various application areas.
2. To identify various key features of Blockchain that will secure artificial intelligence models.
3. To survey various applications of Blockchain for artificial intelligence.
4. To understand how Blockchain will provide reliable and private decentralized data storage for training datasets and integrity of AI models and their predictions/results. The remaining part of the paper is mapped as follows: the study techniques for data collection and extraction and the specifics of data analysis are covered in the second section; the third section provides a discussion on the results of bibliometric analysis with graphical visualization; the fourth section of qualitative analysis outlines the research trends; the article concludes with significant observations, challenges, and future work approaches.

Methodology :

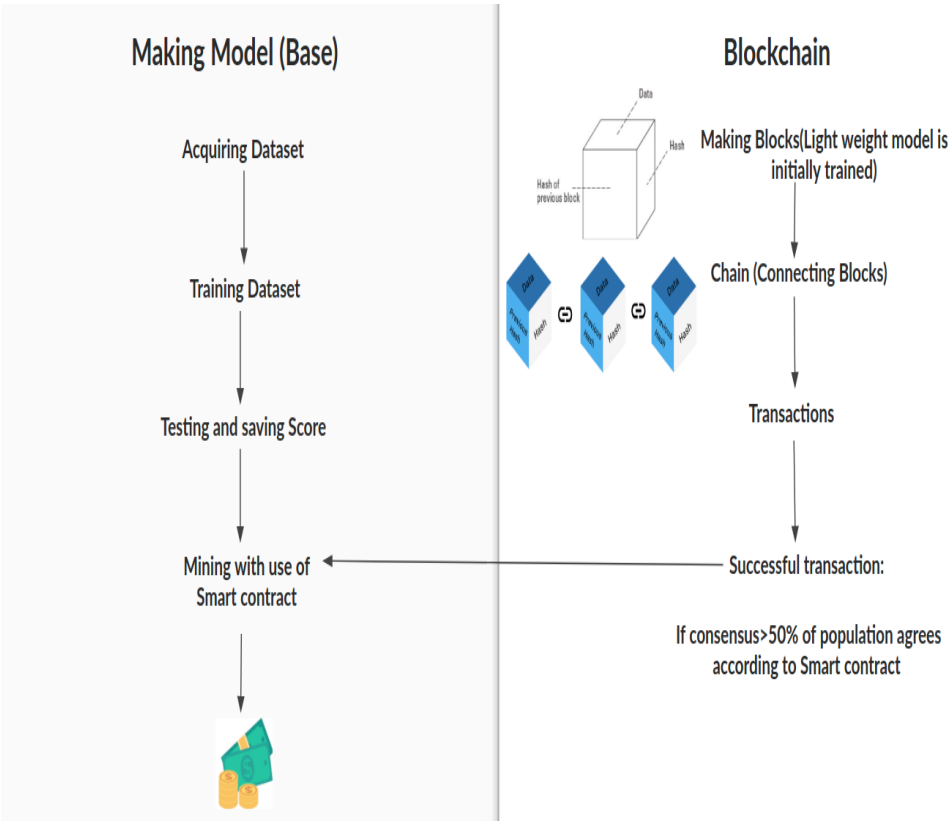


Fig. 4

Chapter -2 Literature Survey

Introduction to ethereum blockchain:

Little is known about peer to peer network for data propagation . Introduction of blockchain gave us exposure for transfer of information securely and without any risk of loss due to redundancy . [5]This enables blockchain consensus .Unlike other blockchains like Bitcoin which is built around one application that is digital currency transactions. Etherium is an open platform where people can build their own applications on top and anything built on top of ethereum gets protected, secured and everything is checked by the entire network .[6]

Distributed Artificial intelligence:

DAI with existing GPVS's is module to module integration with DAI formalization helps us to identify basic common features in the system that is being studied .[7]AI is currently emerging GPVS that could be disruptive in business model across different sectors . [8]
Merging newly emerging technologies like blockchain and artificial intelligence is possible in the current era due to the introduction of cheap storage, GPU and excess processing power that can be utilized in doing complex calculation or a great amount of simple calculation like what happens in GPU for machine learning and deep learning processes.

Database management on blockchain :

The blockchain can sum up to a simple redundant storage system . With the introduction of proof of work and proof of concept it can be used for data storage for AI applications as well as for other usages of redundant secured data.[9] Many medical institutes are using blockchain for keeping medical record safe , the excess processing power can be used for artificial intelligence model to predict the possible disease with already accessed record. In recent research we observed that this can be predict possible breast cancer or diabetic retinopathy upto 4 months before the actual visual symptoms are shown .[9][10]

Current problems faced during AI and database blockchain:

- Privacy : Giving IP address or their data to public network have risk of privacy but due to encryption this can be prevented
- Byzantine attack : Byzantine attack is trusted node in a network going rogue and disrupting the system by providing bad data. This can be solved by proof of work and proof of concept. Even if this situation occurs then with help of version control this can be reversed
- Overwhelming the network [14]
- Spamming : Spamming bad or useless data can also cause problem this can be
- Ambiguous data : In this problem initial data is bad . If the organizer gives bad data then this can cause negatively for this concept as a whole
  [12][13]

Chapter -3 System Development

3.1 Analysis :

DAI requires peer-to-peer support as it is based on blockchain technology .
This project will result in a more reliable model and less computational power . It will result in a huge database and greater knowledge which is usually accessible only to multinational companies like Microsoft , Google etc. Well maintained model which was available for few for application of that model which was available to few organizations will be accessible to more people which will give a better exposure and application .

3.2 Design :

The Design of DAI can be divided into 3 parts :

1) Blockchain
2) Artificial intelligence program (Machine Learning / Deep-learning)
3) Simulation

We made 2 codes for the Artificial program section for Machine Learning and Deep learning respectively and modified our Blockchain section accordingly .

3.2.1 Blockchain :

It is a distributed encrypted database which is shared among various computer nodes for validation , redundancy and security .

- Mother block :
  The initial block which is reference for the models or features that can be added .
- Encryption :
  Traditionally SHA-256 encryption is used with last proof and current proof encoded .
- Register node:

Each newly added computer is given a particular identification i.e. the wallet id for every transaction or reward.

- Blocks :

  Each new block added to the chain contains following:

  - Data:
    - Maker / Profile ID
    - Model
    - Scores
    - Hyperparameters (For Deep learning model)

  - Linking :

    Current block is having id of previous hash with proof of work.
    - Current Hash : Using SHA-256 encryption of universally unique identifiers (UUID).
    - Previous Hash : Hash of previous block

- Smart contract: Proof of work
  - For Model blockchain:
    - The well balanced database is divided into two parts:
      - Mother database: It is the database made for the verification or making the proof of transaction . The model trained is tested on this database to give a better analysis of the model.
      - Child database:

        The child database is made to verify the model by user and to train the model. It is the database that is openly available and users can view and use it without encryption.

  - For Database Blockchain :
    - The Main chain :

Using the scores which are increased or decreased the transaction takes place . The data is added to the buffer chain . To balance the Main chain the buffer chain is transferred to Main chain ,

- The Buffer chain :

  The data is added to the buffer chain depending on the score that is received by the data using the given model. If the score is increased then the data is added after validation .

- Transaction:

  When more than 50% of nodes give consensus according to smart contract the transaction takes place else the transition is rejected .

- Rewarding / Mining :
  - Reward for updation : Similar to bitcoin a minimal amount is given to a node which updates and validates the transaction
  - Reward for Transaction : Here instead of charging a certain amount for a successful transaction . We reward the user for every successful transaction.

3.2.2 Artificial Intelligence Program:

- Database :
  - Score : Using proof of concepts the data is added to the blockchain after preprocessing by trusted nodes . The program first approves the data after running the labeled data through . We used an incentive mechanism to encourage researchers , bounty hunters ,etc. to contribute.
- Model :
  - Score : The score is the correct predictions on the mother database.
  - Machine learning model :

    Using the score we add the model to the blockchain and reward the maker.

    In our project for simulation we used multiple techniques of machine learning on breast cancer detection.

    The following are the machine learning algorithms used :

    - Decision Tree Classifier :

It uses different feature subsets and decision rules at different stages of classification can be classified accordingly . It captures descriptive decision making knowledge from the supplied data.

- Logistic Regression Classifier

  Logistic regression is used when the value of the target variable is categorical in nature. It is most commonly used when the data in question returns binary output(0 or 1).

- Random Forest Classifier

  A data construct applied to machine learning that develops large numbers of random decision trees analyzing sets of variables; helps to enhance the ways that technologies analyze complex data.

- Gaussian Naive Bayesian

  Naive Bayes classifiers are a family of simple "probabilistic classifiers" mainly based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Gaussian Naive Bayes is nothing but a variant of Naive Bayes that follows Gaussian normal distribution and supports continuous data .

- Support Vector Classifier

  Support Vector Machine is one of the most popular Supervised Learning algorithms.It is used for Classification problems in machine learning .The main aim of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future .

- Nearest Neighbors Classifier

  The K Nearest Neighbours algorithm is a supervised learning algorithm. It can be used to solve classification problems. It is an algorithm that makes predictions based on the nature of other data points that are present close to it in the training dataset .

- Deep learning model :

  Using the score we add the model to the blockchain and reward the maker. Here we store the model , layers and hyper parameters used to train the model. For every better score the transaction takes place .

  Here one can use a pre-trained model such as Resnet and customize it to get better accuracy .

  We used Plant disease detection for this project's simulation

  We used following pre-trained models for simulation and one CNN model acting as Mother Block :

  - Resnet50

    ResNet-50 is nothing but a convolutional Neural Network model . ResNet-50 is 50 layers deep. In this you can load a pre trained version of the network which is trained on more than a million images from the ImageNet .(a dataset that has 100,000+ images across 200 different classes)

  - VGG16

```
VGG(
  (features): Sequential(
    (0): Conv2d(3, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (1): ReLU(inplace=True)
    (2): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (3): ReLU(inplace=True)
    (4): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (5): Conv2d(64, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (6): ReLU(inplace=True)
    (7): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (8): ReLU(inplace=True)
    (9): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (10): Conv2d(128, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (11): ReLU(inplace=True)
    (12): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (13): ReLU(inplace=True)
    (14): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (15): ReLU(inplace=True)
    (16): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (17): Conv2d(256, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (18): ReLU(inplace=True)
    (19): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (20): ReLU(inplace=True)
    (21): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (22): ReLU(inplace=True)
    (23): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
    (24): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (25): ReLU(inplace=True)
    (26): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (27): ReLU(inplace=True)
    (28): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (29): ReLU(inplace=True)
    (30): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  )
  (avgpool): AdaptiveAvgPool2d(output_size=(7, 7))
  (classifier): Sequential(
    (0): Linear(in_features=25088, out_features=4096, bias=True)
    (1): ReLU(inplace=True)
    (2): Dropout(p=0.5, inplace=False)
    (3): Linear(in_features=4096, out_features=4096, bias=True)
    (4): ReLU(inplace=True)
    (5): Dropout(p=0.5, inplace=False)
    (6): Linear(in_features=4096, out_features=1000, bias=True)
  )
)
```

Fig. 5

It is a type of convolutional neural network model. This model achieves around 92.7% top-5 test accuracy in ImageNet. VGG-16 is 16 layers deep. You can load a pre-trained version of the network trained on more than a million images from the ImageNet.

- VGG19

  It is also a convolutional neural network model that is 19 layers deep (16 convolution layers, 3 Fully connected layers, 5 MaxPool layers and 1 SoftMax layer).
- Resnet34

  It is a convolutional neural network model. Resnet34 is a 34 layer convolutional neural network. Resnet34 can be utilized as a state of the art image classification model. This is a model that has been pre-trained on the ImageNet dataset .
- DenseNet121

  It is a convolutional Neural Network model which connects each layer to every other layer in a feed-forward fashion. DenseNets have various compelling advantages like they can alleviate the vanishing-gradient problem they can also strengthen feature propagation and not only this they also encourage feature reuse and last but not least substantially reduce the number of parameters

3.2.3 Simulation :

- Machine Learning :

  Validator -> Model selected using score or probability to correctly predicting

- Transaction using best validator :

  Select the best validator and compare to the current transaction after getting consensus using a smart contract .

- Trying to do transaction using Random validator

  Remove the randomly selected part of the blockchain you made and append it to another list.

  For simulation add them in random order to the blockchain after the consensus using smart contract.

- Picking random validator from top performing models.

- Deep learning :

  Making a mother block and trying to do transactions after getting consensus using smart contract .

  - Adding the best model

    Adding choosing the best model and doing a transaction to add the model to blockchain

  - Randomly selecting the model.

    Remove the randomly selected part of the blockchain you made and append it to another list.

    For simulation add them in random order to the blockchain after the consensus using smart contract.

3.2.4 Development

For the application we made a web application using the deep learning model we created and saving the model in the model section in our web application.

With regular auto updation in a server we can always get new best model in regular interval from the blockchain .

The web application is made using HTML, CSS , Javascript for frontend and Flask for server side.

I have used fastai which is built on top of Pytorch. Dataset consists of 38 disease classes from PlantVillage dataset and 1 background class from Stanford's open dataset of

background images DAGS. 80% of the dataset is used for training and 20% for validation.

Frontend:

The frontend uses basic knowledge of HTML , Javascript and CSS

- We have created button and applied an animation for the button
- We created and applied icon for the page
- We applied CSS and the interface of the form.
- Created the basic interface of the page using HTML

Backend:

The backend was created using basic knowledge of Javascript and Flask

- We created the synchronization between Flask (which used the model to provide output) , Javascript (which supported the interface between frontend and backend).
- Created form's backend using javascript and simple error popup system.
- Provided the incoming of image and display of result after going through the model which was trained previously.
- Upload of the image as file.

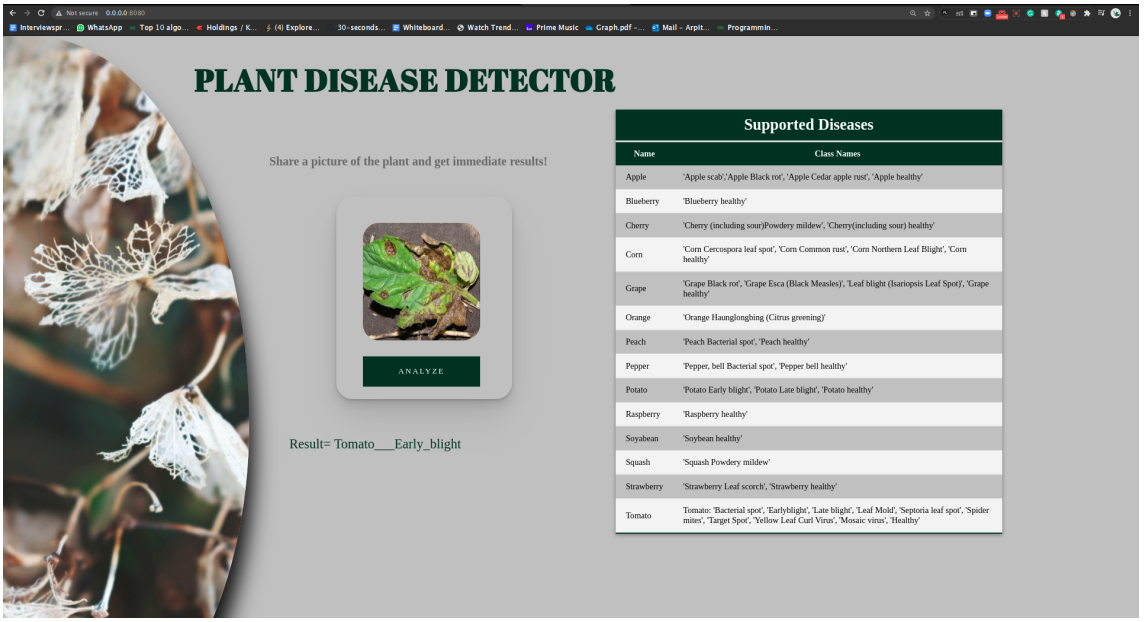


Fig. 6

3.3 Algorithm :

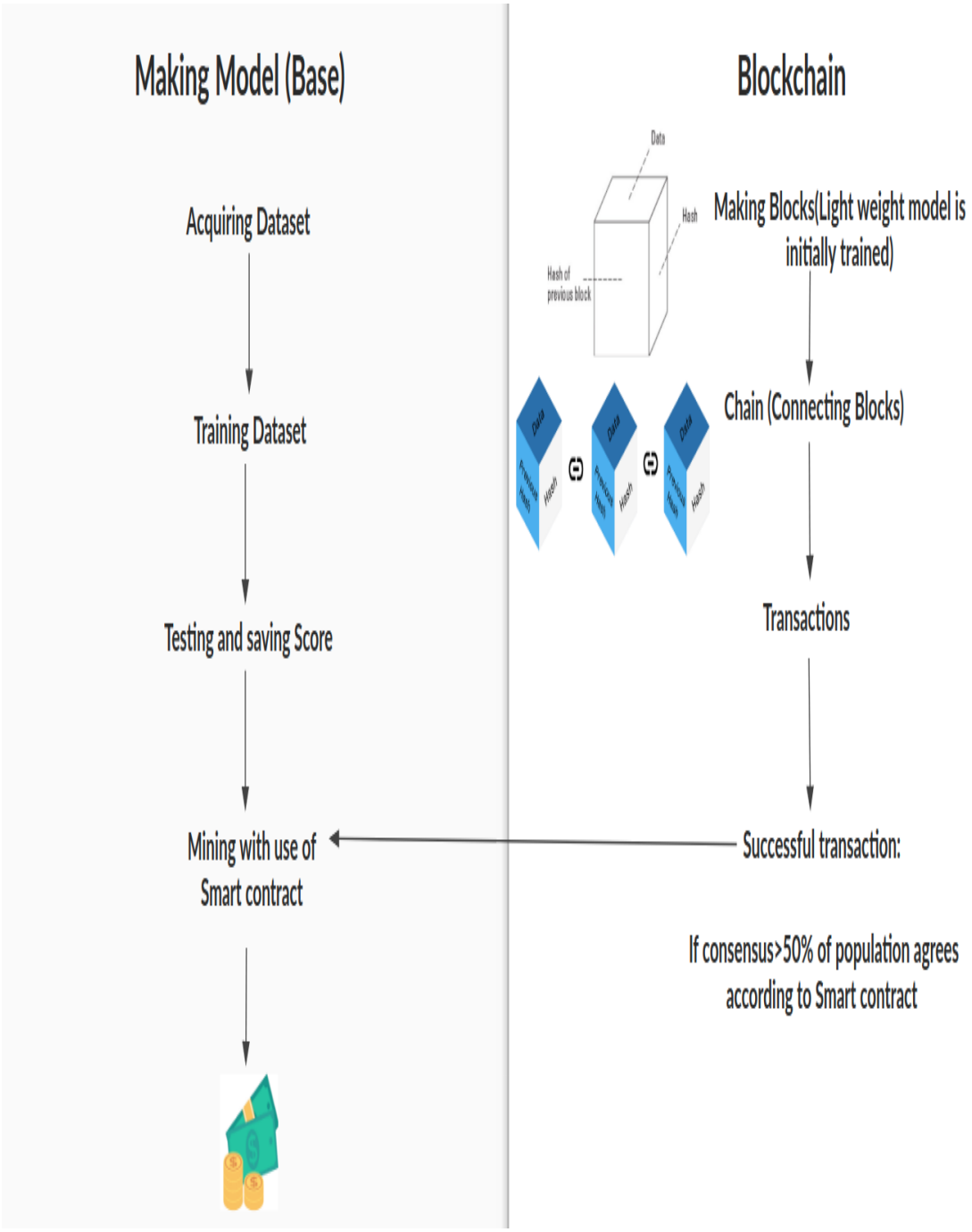


Fig. 7

Machine learning :

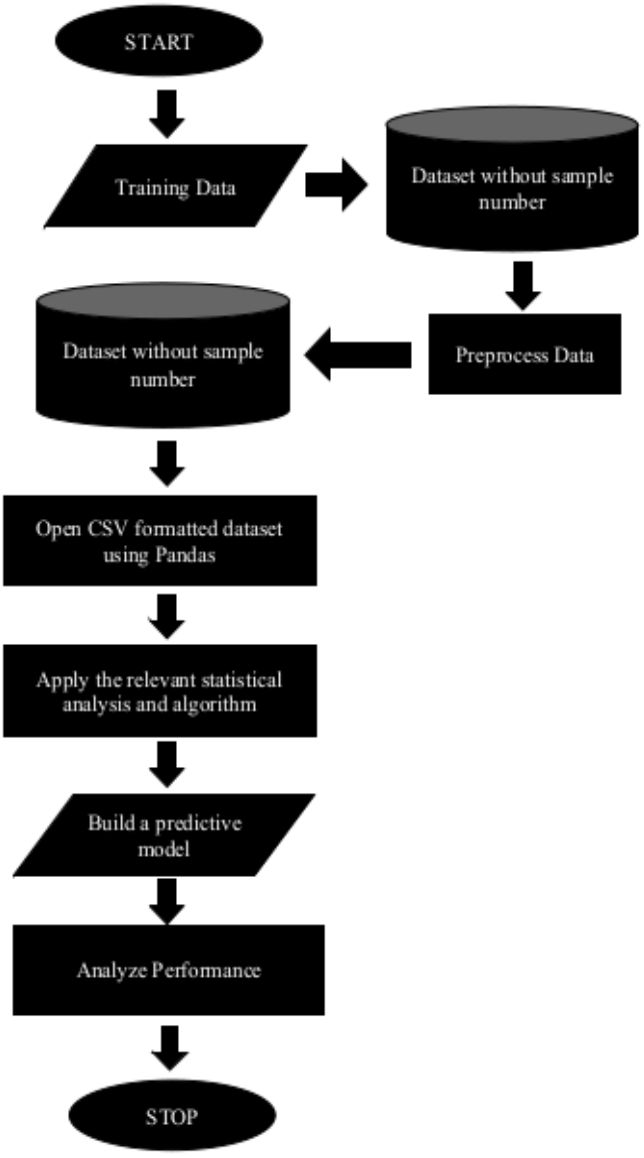


Fig. 8

Commitment Phase:

1. Let A be deposit currency

2. Provider divides the test dataset into 100 equal parts.
3. Provider uploads 100 hashes for each part
4. Smart contract randomly selects 10 of these hashes

Participation phase:

1. for each participants p=1,2….P until number of participants is exhausted or the currency is exhausted:do
2. Participant stakes 1 unit of currency
3. If score provided for given data > score received on best model/validator on blockchain :
4. correct prediction=0 , total prediction=0
5. for each dataset i =1,2...10:
6. for j in dataset part i :
7. if prediction = expected prediction:
8. correct prediction=correct prediction+1
9. total prediction=total prediction+1
10. new score=correct prediction/total prediction
11. if new score>current score :
12. Transaction takes place
13. Reward is distributed

Reward phase:

1. If consensus =1:
2. reward=(new score- previous best score)*Unit of currency

For Database Blockchain :

Making Blockchain:

Chain:

1. Initialize a list chain
2. Create Main Blockchain consist initial database
3. Create Buffer Blockchain consist of different blockchain used to transfer the elements from buffer chain to main chain
    a. MIN_char=inf
    b. for character in characteristics :
    c. MIN_char=minimum(MIN_char,length of character)
    d. Extract MIN_char data from each buffer blockchain and add it to main chain

Block:

1. Initialize a list
2. Set following params :
    a. Index
    b. Timestamp

c. Labeled Data:
d. Consensus 0 or 1
e. Previous hash

Incentive Mechanism:

Commitment Phase:

1. Let A be deposit currency
2. Provider divides the test dataset into 100 equal parts.
3. Provider uploads 100 hashes for each part
4. Smart contract randomly selects 10 of these hashes
5. Defines loss function L(h,D)
6. h->model D->Dataset

Participation phase:

1. for each participants p=1,2….P until number of participants is exhausted or the currency is exhausted:do
2. Participant stakes 1 unit of currency
3. Provides data
4. Save Model updates from h(p-1) to h(p)
5. If L(h(p-1),D)>L(h(p),D):
6. Block is added to Buffer list
7. Reward is distributed

Reward phase:

1. If consensus =1:
2. reward=(new score- previous best score)*Unit of currency

3.4 Mathematical

For best model we received an accuracy of 92.61%

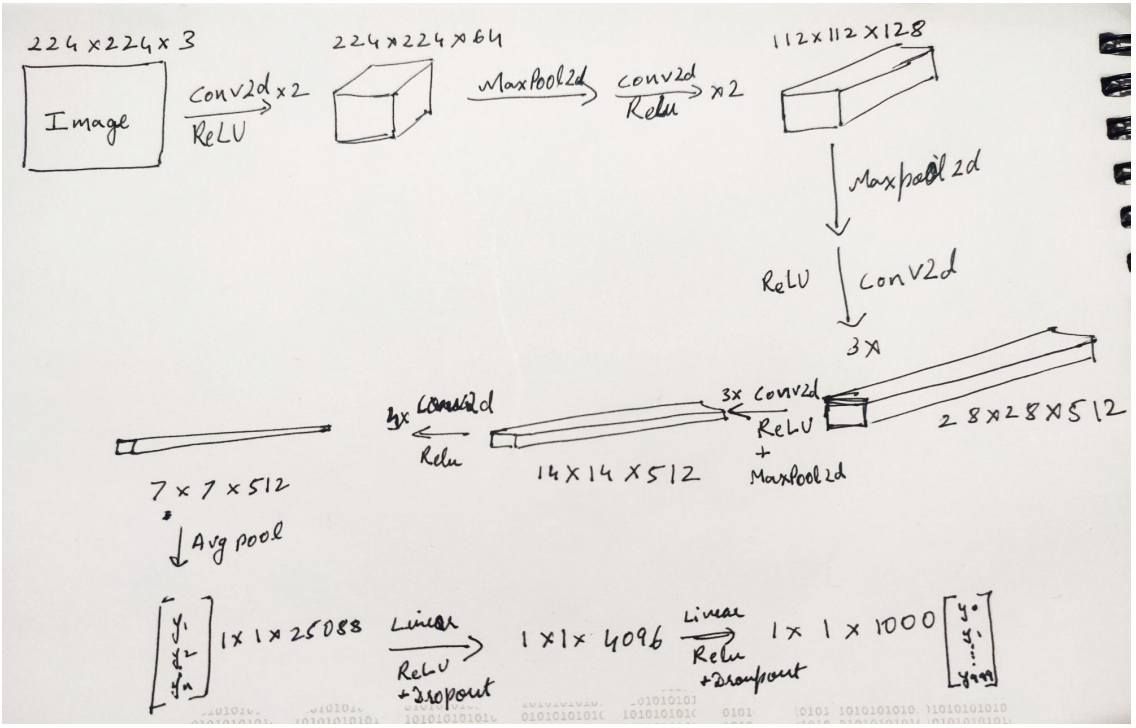


Fig.9

Using the final 1000 features matrix we classify the image.

3.5 Data :

For machine learning model :
The features are taken from a digitized image of a fine needle aspirate of a breast mass .
The data is classified between malignant(Harmful) and benign (Not harmful).
There are 10 real valued features :

a) radius
b) texture (standard deviation of gray-scale values)
c) perimeter
d) area
e) smoothness
f) compactness
g) concavity
h) concave points
i) symmetry
j) fractal dimension

The data consist of 569 subjects out of which 257 are benign and 212 are malignant.

For Deep learning model:

We used a Dataset by SP Mohanty for plant diseases [4] . It is a public dataset of 54306 images for 14 crop species and for 26 diseases . The dataset consists of color ,grayscale and segmented images of the data . The image is resized to 256 × 256 pixels consist of only one leaf

The data is classified into following categories :

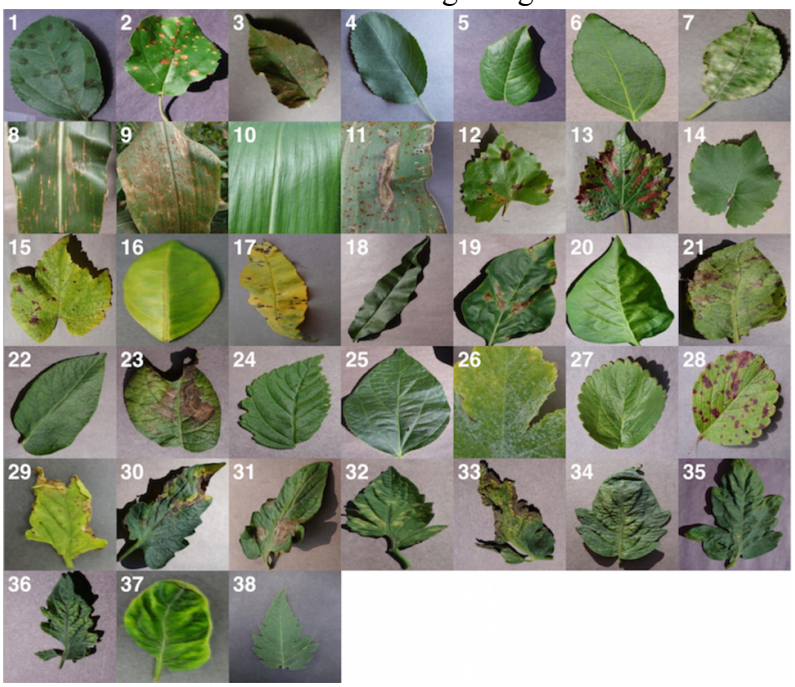


Fig. 10

Which can be sorted according to plant:

Name	Class Names
Apple	'Apple scab','Apple Black rot', 'Apple Cedar apple rust', 'Apple healthy'
Blueberry	'Blueberry healthy'
Cherry	'Cherry (including sour)Powdery mildew', 'Cherry(including sour) healthy'
Corn	'Corn Cercospora leaf spot', 'Corn Common rust', 'Corn Northern Leaf Blight', 'Corn healthy'
Grape	'Grape Black rot', 'Grape Esca (Black Measles)', 'Leaf blight (Isariopsis Leaf Spot)', 'Grape healthy'
Orange	'Orange Haunglongbing (Citrus greening)'
Peach	'Peach Bacterial spot', 'Peach healthy'
Pepper	'Pepper, bell Bacterial spot', 'Pepper bell healthy'
Potato	'Potato Early blight', 'Potato Late blight', 'Potato healthy'
Raspberry	'Raspberry healthy'
Soyabean	'Soybean healthy'
Squash	'Squash Powdery mildew'
Strawberry	'Strawberry Leaf scorch', 'Strawberry healthy'
Tomato	Tomato: 'Bacterial spot', 'Earlyblight', 'Late blight', 'Leaf Mold', 'Septoria leaf spot', 'Spider mites', 'Target Spot', 'Yellow Leaf Curl Virus', 'Mosaic virus', 'Healthy'

Fig. 11

Chapter-4 PERFORMANCE ANALYSIS

4.1 Accuracy For Machine learning :

- Decision Tree Classifier : 92.98%
- Logistic Regression Classifier : 95.9%
- Random Forest Classifier : 96.49%
- Gaussian Naive Bayesian: 92.98%
- Support Vector Classifier : 97.08%
- Nearest Neighbors Classifier: 95.9%

4.2 For Deep learning:

- Resnet50 :

  Accuracy = 89.16%

  F1 Score = 0.8845819234848022

  Confusion Matrix =

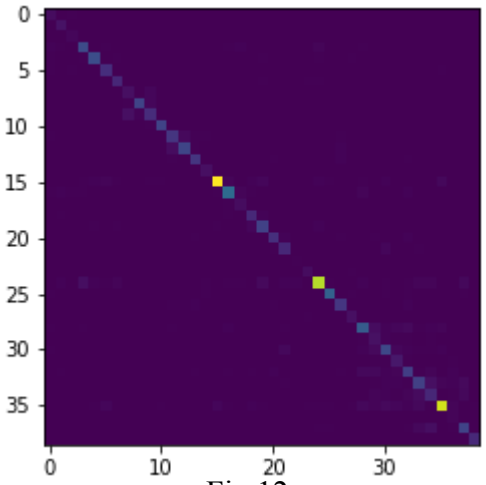


Fig 12

- VGG16 :

  Accuracy = 92.61%

  F1 Score = 0.9922231435775757

  Confusion Matrix =

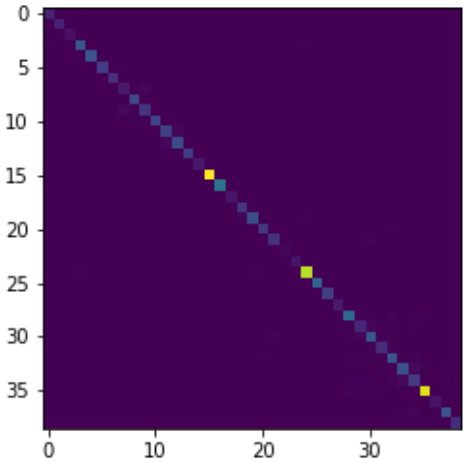


Fig 13

- VGG19 :

  Accuracy = 92.05%

  F1 Score = 0.9936198592185974

  Confusion matrix =

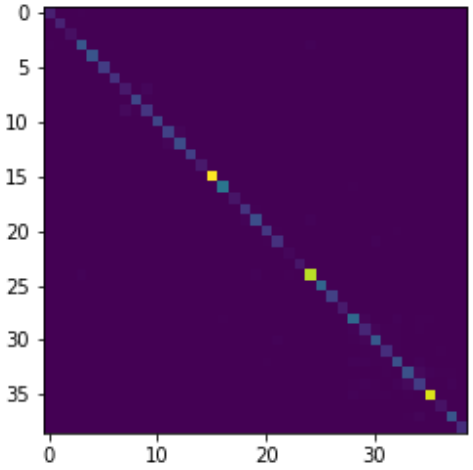


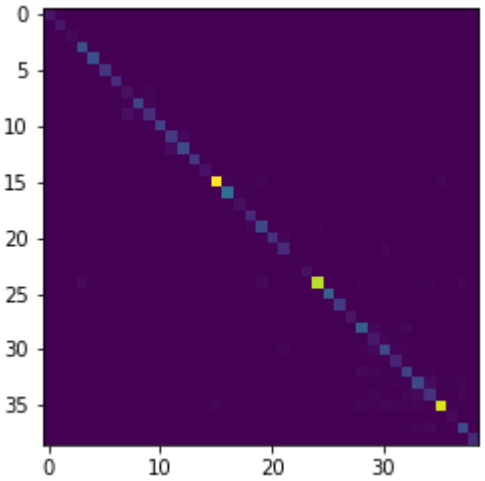
Fig 14

- InceptionV3 :

  Accuracy = 91.62%

  F1 Score =0.9164

  Confusion matrix=



- DenseNet121 :

  Accuracy = 91.15%

  F1 Score = 0.9151003360748291

  Confusion Matrix =

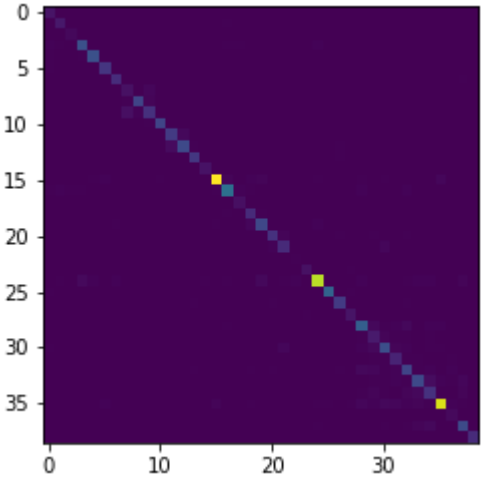


Fig 15

Adding model in blockchain:
Randomly adding best model to blockchain:

- Custom model using LeNet algorithm:

  Accuracy = 94%
  F1 Score =0.94

Layer (type)	Output Shape	Param #
conv2d_6 (Conv2D)	(None, 252, 252, 6)	456
average_pooling2d_6 (AveragePooling2D)	(None, 126, 126, 6)	0
conv2d_7 (Conv2D)	(None, 122, 122, 20)	3020
average_pooling2d_7 (AveragePooling2D)	(None, 61, 61, 20)	0
flatten_3 (Flatten)	(None, 74420)	0
dropout_6 (Dropout)	(None, 74420)	0
dense_9 (Dense)	(None, 120)	8930520
dropout_7 (Dropout)	(None, 120)	0
dense_10 (Dense)	(None, 84)	10164
dense_11 (Dense)	(None, 38)	3230

Total params: 8,947,390
Trainable params: 8,947,390
Non-trainable params: 0

	precision	recall	f1-score	support
0	0.86	0.99	0.92	97
1	0.96	0.89	0.92	123
2	1.00	1.00	1.00	213
3	0.95	0.90	0.92	197
4	0.91	0.97	0.94	220
5	0.99	0.98	0.98	161
6	0.91	0.91	0.91	67
7	1.00	0.95	0.97	893
8	0.96	0.97	0.96	382
9	0.86	0.91	0.89	70
10	0.85	0.96	0.90	139
11	0.90	0.93	0.91	233
12	0.72	1.00	0.84	28
13	0.96	0.84	0.90	174
14	0.85	0.88	0.86	161
15	0.84	0.95	0.89	22
16	0.91	1.00	0.95	40
17	0.99	0.96	0.97	913
18	0.98	0.99	0.98	267
19	0.89	0.99	0.94	168
20	0.96	0.95	0.96	85
21	0.94	0.95	0.94	294
22	0.85	0.84	0.85	148
23	0.88	0.95	0.91	265
24	0.92	0.86	0.89	311
25	0.81	0.94	0.87	125
26	0.88	0.94	0.91	256
27	0.93	0.89	0.91	315
28	0.90	0.86	0.88	222
29	0.98	0.96	0.97	922
30	0.78	0.93	0.85	55
31	0.92	0.93	0.93	223
32	0.88	0.99	0.93	214
33	0.96	0.92	0.94	190
34	0.95	0.92	0.94	137
35	0.95	0.89	0.92	89
36	1.00	0.99	1.00	200
37	0.94	0.94	0.94	132
accuracy			0.94	8751
macro avg	0.91	0.94	0.92	8751
weighted avg	0.94	0.94	0.94	8751

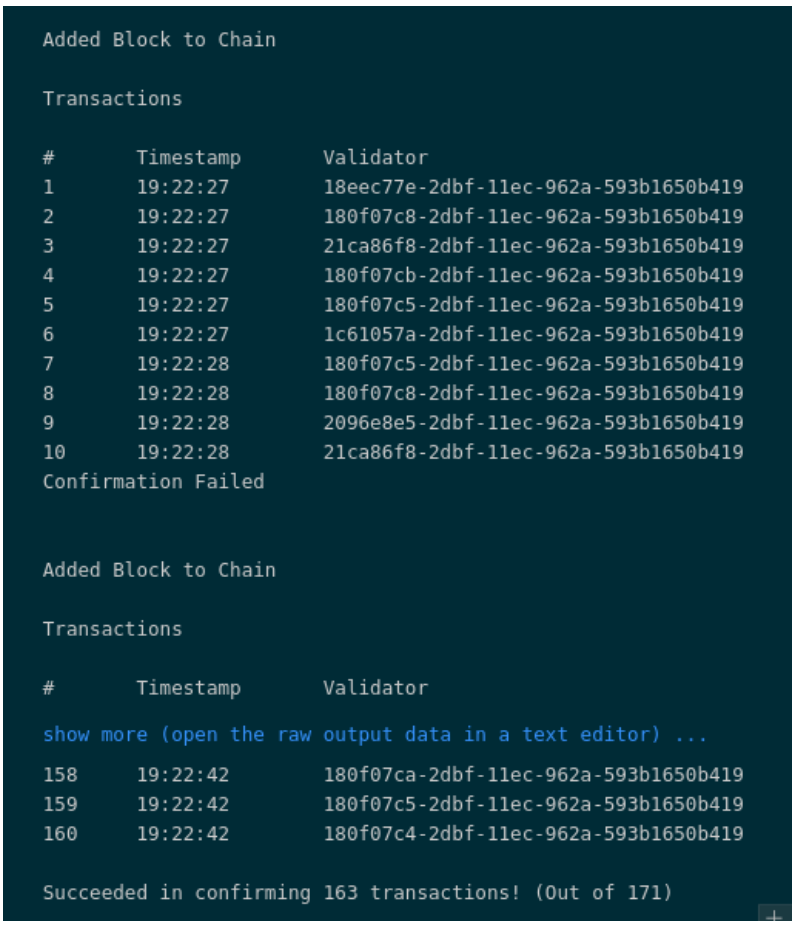


Fig 16

Randomly selecting model and adding them in blockchain for 10000 iteration:

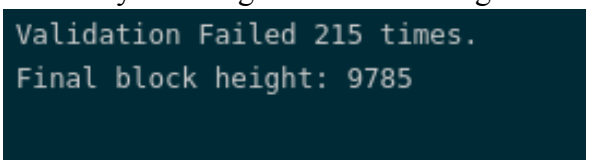


Fig 17

4.3 For Database Blockchain :

We addressed the problem or flaws stated by Microsoft in there recent publication of paper and tried to improve the smart contract for solving the issue.
Issue stated: Bad data insertion -> Bad data can be introduced by a trusted node by showing a better result but could be miscellaneous in nature .[1][2]
Solution: We proposed using proof of work as well as proof of concept. We randomly selected 10 parts of data or 10% and used it for merging and comparing it on the 90% data . After proof of work for each transaction proof of concept is introduced for proper data insertion.

5. Chapter-5 CONCLUSIONS

The achievement of my current thesis is researching the future perspective of blockchain technology and artificial intelligence and how it can be implemented and used to create various applications to assist the economy, developers and researchers to provide efficient solutions .

In the application section we saw a web application that used model to predict the data , with auto updating this can automate the process of development and provide an efficient model. Our research highly depends on peer to peer network for updation of data and model .
The data updation system still needs some improvements and it still has vulnerabilities when low data size is used ,which further needs some research to make it more sustainable.

5.1 Outcome :

We created Decentralized Artificial Intelligence or DAI on a blockchain (on one system for simulation).
During phase 1 we made machine learning models on breast cancer detection using various algorithms and compared their performance, getting a best accuracy of 97.08% . We were able to successfully run a simulation by making a blockchain .For phase 2 we updated the blockchain including hashing , encryption and smart contract . We made a rewarding algorithm for the user .
We made Plant disease detection model using various algorithms, achieving an accuracy of 92.61% and made a simple user interface for the system. We tried to auto update the best model from the blockchain to the web page.
We also propose an algorithm for database management for the AI model in blockchain.

5.2 Future Scope :

Using public blockchain:

Implementation of DAI in Ethereum like blockchain . Ethereum is a general purpose blockchain . Unlike other blockchains like Bitcoin which is built around one application that is digital currency transactions. Etherium is an open platform where people can build their own applications on top and anything built on top of ethereum gets protected, secured and everything is checked by the entire network .

Contribution of professionals and participation of developers:

The objective of this project is provide developers a superior model that they can make , so a person with better knowledge of hyperparameter tuning and with the introduction of new techniques can update the model/data for developers to use and thus creating better

and more creative applications. The updation of model and data will continue and over time reach its full potential ,and thus shifting knowledge , data and power of development from big organizations to small/ individual to develop applications on par with them. People who do not have a gpu or much budget to develop models can also use models.

Faster transactions
Since blockchain is decentralized and peer to peer connected , for sharing a file or for transaction a group of computer nodes is required . The peer to peer communication happens between Multiple Client Devices .
In future , the chain or data can be split into “Bits” as a peer wants to download the data . all the devices can simultaneously upload and download the data and thus increase the speed of transactions happening .

Models:
More research can be done on determining which models are more compatible with this system and reduce the size of blockchain. There is cost limitation with data updation and model updation . We propose more research in pre computing as much as possible in a computer rather than on blockchain to reduce the cost and only performing necessary steps.
Use of transfer learning can be used for reducing the size of blockchain .

We are aiming for updation for a model instead of replacement for reducing the overall cost for a more complex model.

REFERENCES

[1] Harris, Justin D., and Bo Waggoner. "Decentralized and collaborative AI on blockchain." 2019 IEEE International Conference on Blockchain (Blockchain). IEEE, 2019.

[2] Harris, Justin D. "Analysis of Models for Decentralized and Collaborative AI on Blockchain." International Conference on Blockchain. Springer, Cham, 2020.

[3]Li, Guanbin, and Yizhou Yu. "Contrast-oriented deep neural networks for salient object detection." IEEE transactions on neural networks and learning systems 29.12 (2018): 6038-6051.

[4]Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using deep learning for image-based plant disease detection." Frontiers in plant science 7 (2016): 1419.

[5]Tikhomirov, Sergei. "Ethereum: state of knowledge and research perspectives." International Symposium on Foundations and Practice of Security. Springer, Cham, 2017.

[6] Kim, Seoung Kyun, et al. "Measuring ethereum network peers." Proceedings of the Internet Measurement Conference 2018. 2018.

[7] Boissier, Olivier, and Yves Demazeau. "A distributed artificial intelligence view on general purpose vision systems." ACM SIGOIS Bulletin 13.3 (1992): 16.

[8] Delponte, Laura, and G. TAMBURRINI. European Artificial Intelligence (AI) leadership, the path for an integrated vision. European Parliament, 2018.

[9] Zheng, Xiaoying, Yongxin Zhu, and Xueming Si. "A survey on challenges and progresses in blockchain technologies: A performance and security perspective." Applied Sciences 9.22 (2019): 4731.

[10] Qummar, Sehrish, et al. "A deep learning ensemble approach for diabetic retinopathy detection." IEEE Access 7 (2019): 150530-150539.

[11] Baker, Larry H. "Breast cancer detection demonstration project: Five‐year summary report." CA: a cancer journal for clinicians 32.4 (1982): 194-225.

[12]A. Marathe, K. Narayanan, A. Gupta, and M. Pr, “DInEMMo: Decentralized incentivization for enterprise marketplace models,” 2018 IEEE 25th International Conference on High Performance Computing Workshops (HiPCW), pp. 95–100, 2018.

[13]Harris, Justin D. "Analysis of Models for Decentralized and Collaborative AI on Blockchain." International Conference on Blockchain. Springer, Cham, 2020.

[14] Alsamhi, Saeed H., et al. "Blockchain for decentralized multi-drone to combat covid-19." arXiv preprint arXiv:2102.00969 (2021).

APPENDICES

Referred data :

Plant disease detection by SP Mohanty .

Breast cancer detection

Referred chart:

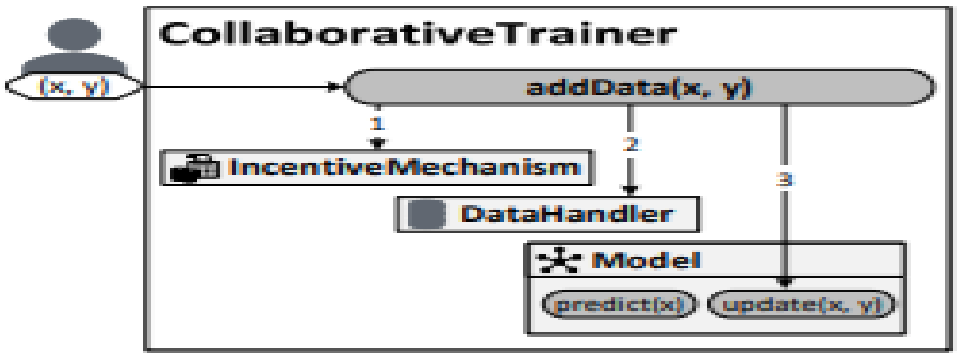


Fig.9

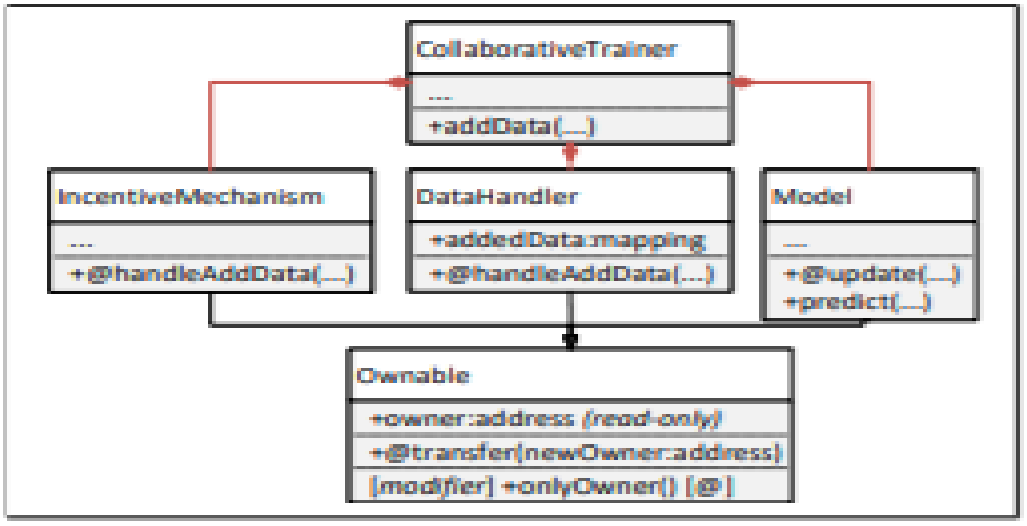


Fig. 10