Quantification and Mitigation of Epidemic using Deep-Learning on Distributed Peer to Peer Network

Shresth Agrawal
Guide- Sushil Kumar Pandey
shresthagrawal.31@gmail.com | sushilpandey0176@gmail.com

Abstract- *Contagious Diseases are the deadliest killers, and there exists no pre-eminent approaches to predict, quantify and mitigate them. Existing mathematical models miserably fail on the real world due to their rigidity towards a specific kind of diseases and poor data collection. The scale of problem vs the organizational working capacity creates a bottleneck problem. The intensity of the data collections requirement may lead to privacy and data breach problems. I propose a novel solution to this problem by using Neural Networks on Ethereum Block Chain. Neural Networks are excellent at dimensionality reduction and supervised prediction which makes the model more flexible. First, the data is uploaded in real-time through a thin client of web3.js to the IPFS hash tables, the generated hash addresses are updated in the smart contract of the blockchain, this smart contract consists a pre-defined neural network and the generated hash address of the training data. The blockchain miners use this data to train a neural network as a 'proof of work' and the results will again be stored in the IPFS hash tables. The user can query the blockchain to get the IPFS hash address where the results of the neural network have been stored and in return will pay crypto-tokens (currently ether) to incentivize the miner and data uploader. This leads to a completely autonomous and immutable system for epidemic management. It is elastic as the system grows with the increase in problems, and secure as the whole system works on zero trusts and only cryptographic security.*

I. INTRODUCTION

"Contagious diseases is the world's biggest killer which kills approximately 15 million people annually"- **WHO**. There are several mathematical modelling on Epidemics. Mathematicians have generally used the SIR model for forecasting epidemic, it was first proposed by A.G. Kendrick and W.O. Kermack in 1927, in the research paper- "A CONTRIBUTION TO THE MATHEMATICAL THEORY OF EPIDEMICS', where a simple deterministic model was formulated which was successful in predicting the behavior of the outbreak. In 2017, I drafted a project "**A Real-Time Predictive Modelling for Mitigating Contagious Diseases**", (which was also a part of ISEF 2017), where I tried to improve SIR model by including various factors and created a real time application for diseases surveillance and management. Till date, each and every algorithm including my work, has certain limitations:

- They generally display less efficiency on real world, due to their rigidity of predicting a specific type of disease, or by including a certain set of factors.
- Minimal or a poor data collection method.
- The Bottle-Neck Problem: Diseases are growing exponentially but the organizations working for their cure are small.
- Modeling factors like migration may also require intensive data collection through some organizations which may lead to data breach & security problems.

Hypothesis: By creating a Neural Network based system on Decentralized peer to peer network can create a scalable, autonomous, flexible and secure system for epidemic management.

Objective: Deep learning, which could be implied on any form of data, could integrate flexibility to the model as compare to the existing SIR based models, which have rigid factor specifications. Creating the algorithm decentralized, would make it more scalable with the spread of disease, the nodes in the network will autonomously increase, which will automatically lead to increase in system size, autonomous as everything will be based on smart contracts and no central authority can govern this, and also secure as everything will be based on cryptographic proofs.

II. LIME-LIGHT / NOVELTY

There exists several mathematical modeling for predicting epidemic but this project introduces a completely novel approach by using neural network on decentralized peer to peer network to create a system quantify epidemic. This system uses real time data collection which is securely stored on an IPFS (Interplanetary File System), and uses smart contracts on ethereum blockchain to create a complete autonomous system, which is immune to any hacks or censorship.

III. ALGORITHM

Data Collection Phase: first the data is collected on real-time through a web-app which uses web3.js and IPFS http API to connect to Ethereum and IPFS. Data can be submitted by hospitals, health care centers and Infected people themselves. The web-app itself is stored in the IPFS itself and can accessed through an initial hash address. Anyone submitting the data needs to generate a public private key use elliptic curve digital signatures which will be further used to incentivize them with crypto-token (ethers). For initial research we limit the data parameters to the number of infected, number of removed (recovered or dead), the current time stamp and the current coordinate location. For every patient that gets registered we will take their personal details (name, age and address) and hash it with SHA256 to generate a unique id and will only store the unique id in the IPFS, this will avoid overlapping of same data to be entered by different users.

Mining /Training the Neural Network: As soon as a certain data requirement is fulfilled a smart contract programmed using Solidity will be added to the ethereum block chain consisting of the specified architecture of Neural Network and hash address for training data set. The neural network consist of five nodes in the input layer, then there is one hidden layer whose number of nodes is dynamic and the output layer consist of two nodes. The neural network uses tensor flow lite which removes the requirement of floating point in the blockchain as well as decreases the computation required. The miners have to download the data as well as the code, train it locally and put the results back in the IPFS hash table, this would be used as 'Proof of Work' to tokenize the miners.

End User: During an outbreak when a user is suspicious to know about their probability of infection or government requires a furcating of the outbreak to create specific policies for mitigation they can pay some crypto token using the same Web-App which will query the blockchain to get the hash address of the results.

IV. METHODOLOGY

Testing Neural Network: After coming from ISEF2017 I started piloting my project, during that time I realized that for every other disease I had to model a different equations which made my project cumbersome and slow on real time, The system of linear differential equation that I used in my previous project were not integrable and the only way to approximately get the answer was to do Numerical Integration which decreased the accuracy significantly. After continuously testing with different statistical approaches I found Neural Network to be the most efficient as now I didn't need to remodel equations every-time and accuracy was also not hampered. I used the IDSP (Indian Disease Surveillance Program) data, and EBOLA data from WHO to train and test the Neural Network and results were significantly higher than the previous project.

Decentralization and Use of Ethereum: Using a centralized system could lead to various problems, first it is prone to DDOS (Denial of service) attack, second during an outbreak a single server will contain lot of intensive data if a hacker somehow gets the access this may lead to huge data breach problems, third while there are several outbreaks simultaneously this would slow down the server and hence has scalability problems solution to all the above problem is using Decentralized system where the system works autonomously. Ethereum is perfect for this task as it allows users to write custom code snippets (smart contracts) using solidity in the block chain and also provides the consensus layer to the DAPP. I created a webapp using JavaScript which used web3.js to communicate with the ethereum and test it I used Test-Rpc and Truffle as blockchain simulation.

Storage and Computation Problem: storing all the data collected in the ethereum blockchain would again lead to scalability issues in long run, hence I decided use IPFS for file storage, in IPFS the file was divided in small chunks of data and hashed the hash is then used to access the data. To add files in the IPFS using the webapp I used IPFS HTTP API. The webapp itself was added to the IPFS which made even furthermore decentralized. The next issue was computation, adding Neural Network to the blockchain with normal Tensorflow backend would require floating point which was not available in the ethereum blockchain hence I used Tensorflow lite.

Data Overlapping Problem: During data collection it is possible that two different user might submit the data of same patient which would cause data inflation, and if a database of details of the patient is kept and matched every time a new patient is registered this would also lead to Privacy issues hence I decided to take the personal details of the patient JSON encode it and only store the hash (SHA256) of it which would remove the privacy issue and also solve the overlapping problem.

V. Observation and Conclusion

The above system clearly introduces a novel approach for creating an autonomous, scalable, secure, flexible approach in quantifying the outbreak of contagious disease using a Deep learning in distributed peer to peer network.

Link to Campaign: https://salabs.in

V. REFERENCE

[1]. A Real-Time Predictive Modelling for Mitigating Contagious Diseases
https://drive.google.com/file/d/10pIBJjZKdbUIKGx9XW0mN6N7A6Kh8fSL/view

[2]. A Proposal for Epidemic Prediction using Deep Learning
https://www.technoarete.org/common_abstract/pdf/IJERCSE/v4/i3/Ext_01563.pdf

[3]. A Next-Generation Smart Contract and Decentralized Application Platform
https://github.com/ethereum/wiki/wiki/White-Paper
[4]. Trustless Machine Learning Contracts; Evaluating and Exchanging
Machine Learning Models on the Ethereum Blockchain
https://arxiv.org/pdf/1802.10185.pdf
[5]. Bitcoin: A Peer-to-Peer Electronic Cash System
https://bitcoin.org/bitcoin.pdf
[6]. www.ncbi.nlm.nih.gov/pmc/articles/PMC3162588/
[7]. www.sciencedirect.com/science/articles/pii/S1198743X14604080
[8]. Mathematical Biology, J.D. Murray, Springer-Verlag, 1989
[9]. Using Real Data in an SIR Model D. Sulsky
[10]. https://www.cdc.gov/about/facts/cdcfastfacts/contagious-diseases-disasters.html
[11]. https://sites.math.washington.edu/~morrow/mcm/mcm15/38725paper.pdf
[12]. https://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/index.html?s_cid=cs_4926
[13]. *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*
[14]. http://web.iitd.ac.in/~pmvs/courses/mel705/curvefitting.pdf
[15]. http://rspb.royalsocietypublishing.org/content/267/1458/2219.short
[16]. https://search.informit.com.au/browseJournalTitle;res=IELAPA;issn=1447-4514
[17]. http://www.sciencedirect.com/science/article/pii/S0895717700000406