D-App for providing geriatric aid and to collect and handle periodic health survey data

Chinnu Joseph\textsuperscript{1,2}, Chinchu Joseph\textsuperscript{1,3}, Atul Krishna P\textsuperscript{1,4}, Amal George Mathew\textsuperscript{1,5} and Fabeela Ali Rawther\textsuperscript{1,6}

\textsuperscript{1} Department of Computer Science and Engineering, Amal Jyothi College of Engineering, Kerala, India
\textsuperscript{2} chinnujoseph29@gmail.com
\textsuperscript{3} chinchujoseph2020@cs.ajce.in
\textsuperscript{4} atulkrishnap2020@cs.ajce.in
\textsuperscript{5} amalgeorgemathew2020@cs.ajce.in
\textsuperscript{6} fabeelaalirawther@amaljyothi.ac.in

Abstract. The system aims to build a health survey platform designed to collect basic health information about residents of particular area. Block chain is used to store the collected information making sure only authenticated persons can have access to crucial information. This system also provides information about government health policies or schemes that the members of a family are eligible for and can have easy access to these facilities. This system can also be used as a platform to host health survey in case of an unexpected outbreak of a disease. Machine learning algorithms may be used predict health issues like the possible occurrence of epidemics etc. The system efficiently reduces the paperwork and inefficiency of traditional door to door health surveys by enabling each family to directly feed in their information.

1. Introduction

Machine learning is one of the most thriving technologies that the world has come across. This technology is still in its developing stage. It will undoubtedly continue to attract more people in the coming years as well as make headlines. Machine learning is a tool for turning information into intelligence. In recent decades, there has been an acceleration of data. Such data are worthless until we study them and extract the pattern within them. Machine learning is used to identify useful fundamental correlations efficiently within complicated data that we might otherwise fail to discover. Secret trends and question information can be used for potential activities and for sensitive decisions of any kind [1]. The use of a shared cloud consequently leads to security problems in data storage and processing. So, we need some sort of security algorithm that grants a scalable, key-based encryption system that can resolve delays, security forces, and storage threats.

The history of any digital ledger is irreversible and transparent, through the use of decentralization as well as cryptographic hashing. While generating the first block of a chain, nonce produces the cryptographic hash. The data in the block is characterized as signed and linked forever to the nonce and hash unless it is extracted. Each block comes with its distinct nonce and hash but also relates to the hash of the previous block in the chain which makes it difficult to mine a block particularly on big chains. Decentralization being the most significant principle of block chain, no one machine or organization can own the chain [1]. Every node has its block chain replica, and any freshly mined block requires the
network to update, authorize, and test the chain algorithmically. Since these distributed ledgers are transparent, any change in the ledger can be quickly checked and viewed. A specific ID number representing the transaction is given to each user which is alphanumeric. The fusion of digital records with a network of checks and balances aims to preserve transparency and create consumer trust [3].

As machine learning and block chain converge, people will benefit from ML’s ability to automate the handling of a massive amount of data. In fact, it would be possible to create a completely new paradigm by putting the two together. By using ML to manage the chain, there is also an opportunity to drastically improve security. In addition, since ML enjoys working with a lot of data, this offers the opportunity to develop better models by utilizing the decentralized nature of block chain which also promotes data sharing. Sometimes you might end up with a qualitatively new data set that is also a better data set when all the data from the silos converge [3]. As a result, it will lead to the creation of a qualitatively new model where new insights can be derived which, in turn, can provide new opportunities to build cutting-edge business applications for next-gen. This can be a game-changer for the majority of industries as it can be used as a tool for fraud detection [1]. Thanks to a common ledger framework.

2. Proposed System

This research was initiated as a way to increase public awareness of the numerous government services and initiatives for the welfare of the elderly [4]. We wondered why not develop apps to bring the specifics of the organizations and student bodies operating health camps and inspections to the geriatric crowd. Further brainstorming led us to a study network idea that allows everybody to be involved in their health care [5]. This platform can provide the public with strong support for the government in relation to health issues and in turn, helps the ministry study people’s general health. The second reason for this idea is that the government is not prepared to have an electronic survey forum during an infectious disease epidemic, and it would be of great benefit in our country to create a website that could implement all the above criteria in a limited period [6, 7].

The government struggles to gather its people’s data in an incredibly populated country like India [4]. The gathering of data proceeds in conventional formats, e.g. door-to-door surveys. Only with regular and accurate surveys can the government analyze the health, needs, and statistics of its people [8]. The major issues that should be taken into consideration for improved surveys for keeping track of trends in India’s disease burden are the inability to collect data on non-communicable disease, injuries and the interval between the survey rounds, another area that needs attention is the delay in making the data available in the public domain [9]. Currently, there is no notification system for the public. We often do not know many of our Government’s incentives, which lead to a lack of efficiency in many of our plans. We do not have a platform for an online survey to gain data to study it for medical conclusions even following an outbreak of a disease or a pandemic. These issues should be solved by innovations in data communication and processing technology. The latest computational methods and tools for machine learning are used to effectively enhance our country's health quality by providing an online health database.

The whole network is running on a block chain network, which is an advanced data protection system commonly used. The data storage is unlike the relationship database but as a chain, i.e. one connects to the other; all new data or any data updates made are added as a next block in the chain. Every activity in the network is logged in and every attempt at unauthorized access is strictly controlled by block chain technology [10].

Our system uses an open-source block chain. It provides smart contract functionality, decentralized block chain. Now, the smart contract is a set of rules to be followed by all participants in the block chain network [11]. It is this contract that enables us to specify what actions need to be carried out and who all are authorized to do so. The contract is deployed as the first transaction in the block chain and therefore each node has the same version of the code. We used the language of solidity programming [12] to write our smart contracts, the language has features from C++, python, and JavaScript and is specifically designed for the Ethereum virtual machine.
Truffle development [13] and testing framework have been used for our prototype. This framework allows you to compile and deploy your smart contract on a blockchain network. It also provides testing for the behavior of your contracts and thus saves you from unexpected errors after network setup. Since the aim was to build a working prototype of the proposed system, we have used Ganache blockchain, which is a personal blockchain for the rapid development of Ethereum applications. Ten blockchain accounts are provided so that we can develop, deploy, and test our Dapp in a safe environment. You can easily connect the truffle network to the Ganache. Therefore, we are freed from having to set up nodes in our network.

Web3.js is a library set that will communicate with the blockchain using the HTTP-connection (this server is operated by a Ganache). It will access user accounts, send transactions; interact with smart contracts, and more. Extension Meta mask in our browser permits Dapp blockchain access, propose transacting, and reads user account address. Every transaction must be checked by the account owner before being executed.

The user can access the website through their blockchain account via the Meta mask. If the account is already registered the account holder profile is displayed, otherwise registration. They are provided with a registration page where they register themselves under their blockchain account. They have to specify their age, which lets the system separate its users into geriatric and non-geriatric. Once the users are registered, they can view the incentives or schemes they are eligible for and approach the ministry to avail it. If the government sets up a survey, they can give their responses by logging into the system. In case you develop a new disease, you can update the same on your profile. The user can also view a general analysis of the situation in case of a pandemic or epidemic thus staying updated with the correct information.

On the other hand, the government account holder can view all the details of the system users whenever required. Another main responsibility of the government account is to add incentives and schemes and their related information. These are then matched to all the eligible users of the system. The
government also gets to see the survey results and thus get an overall analysis of the major diseases among its citizens.

This system can also be used as a host health survey in the event of an unexpected disease outbreak. Machine learning is used to analyse and link symptoms to illnesses. N number of symptoms and n number of illnesses are fed to the machine to train the model. The model is tested to reach the required accuracy. When the user completes the survey, the symptom survey response is registered and mapped to the disease using the trained and tested machine learning algorithm. The output of this study is the likelihood of diagnosing a particular disease to a given person. All the data and results from this step are stored in the block chain. Machine learning algorithms can be used to predict health issues such as future outbreaks (fig 1).

3. Experimental Results
The predominant objective of this system is to perform a survey among a large number of people and to derive an analysis of the data that may be useful to the government [2]. With the help of a data set of 132 symptoms and 41 diseases commonly occurring around a year, we have mapped out a predicted disease after the completion of the survey. The algorithms used for the designed system are decision tree and KNN [2]. In our system, symptoms that give the better classification are assigned higher importance and are thus concentrated near the root node, and other towards the leaf nodes of the tree, providing a more accurate prediction based on the more prominent symptom related to that disease. As already known, the output of a decision tree is a single class, thus our system accurately predicts one of the prominent diseases from those symptoms provided by the user, but chances of presence of multiple diseases to be detected from a survey is also high, thus we need another algorithm which can detect a less evident disease also. K Nearest Neighbour (KNN) is used complementary to the decision tree, in our system. KNN is the most effective pattern detection lazy classification algorithm [2]. The closer neighbours contribute more to the resulting class than the more far off ones. The number of neighbours which are assigned is 10 for the developed system. With even the lower number of neighbours, we can detect a less evident disease from the survey. Since the platform surveys are taken at regular periods the presence of more than two undetected diseases in a person is very less likely to occur. Given four-five prominent symptoms of a disease, we can achieve a prediction with high accuracy from our system. The combination of both the algorithms effectively provides disease that is more obvious along with comparatively less obvious. Hence combining the predictions from the decision tree and KNN we get an effective result of predicted diseases. The table snippet (fig. 2) shows the prediction results from a sample of 34 survey responses. This predicted disease of each individual for a survey is stored under a common serial number in the block chain. The government can view a graphical representation of the count of diseases detected in a region for that survey, thus aiding them to take timely action.

The time series prediction of valuable data using Long Short-Term Memory (LSTM) networks is another component of this system where power machine learning is deployed. During times of a pandemic, like COVID-19, it is the key duty of the government to provide accurate data of the number of confirmed cases on a daily basis to its citizens. Along with this, the government should be able to get an approximation of number cases that might develop in the coming days. Our platform has developed a model that predicts the total number of confirmed cases from all the countries with the presence of COVID-19, up to 10 days with approximate range accuracy (fig. 3).
The data set used for the prediction was compiled by the Johns Hopkins University Centre for Systems Science and Engineering (JHU CSSE), supported by ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL) [14]. The model was trained with 15 hidden layers and 1500 epochs to achieve the required accuracy over a test data of 60 days. The available and predicted data are graphically displayed to provide the user with improved interpretation or analysis. The LSTM can learn long-term dependencies without the problem of the vanishing gradient, which affects their performances, and can tackle sequential and temporal problems. LSTM has a chain-like structure, an RNN equivalent, except there’s four, composed of a cell, input gate and output gate and a forgot gate consisting of Sigmoid and Tanh functions, rather than providing a single layer of the neural network.

### 4. Results

The system prototype shows a decentralized web application that will allow every citizen, distinctively the elder citizens, to be aware of the government-funded special monetary and nonmonetary medical schemes and check-up activities [4]. We successfully mapped the right health incentives (which are made available by the govt and NGO’s) to the elderly that requires it using solidity datatypes. The survey

| DISEASES         | COUNT |
|------------------|-------|
| Pneumonia infection | 1     |
| Allergy          | 1     |
| GEMD             | 1     |
| Chronic mycosis   | 1     |
| Drug Reaction    | 1     |
| Peptic ulcer disease | 4     |
| AIDS             | 0     |
| Diabetes         | 0     |
| Cardiovascular   | 1     |
| Renal failure     | 0     |
| Hypertension     | 3     |
| Malaria          | 0     |

**Figure 2** The table shows the list of diseases and the corresponding no. of people predicted to be having those diseases. Every time the particular survey result, predicts a disease, the count is incremented; thus, giving the government quick analysis of the trending disease.

Figure 3 The figure shows the plot of historical daily COVID-19 cases used for the training of LSTM model along with comparative plot of predicted number of cases and the test data of real number of cases.
page allows easier access to mass health information in short time notice. As described in the experimental section above, ML algorithms were used to process the collected survey data to give us insight into the predominant and trending disease among the citizens.

5. Conclusion
The website will make the annual health survey digital and provide a much easier way to analyse the survey responses. Health surveys can be conducted in shorter intervals with minimum human resources and government authorities responsible for conducting them will become more answerable about the task assigned to them. It will greatly aid timely action on healthcare. Another aspect of this system involves mapping the diseases of an individual to available incentives from the government or other NGOs especially for the elder members of the society enabling them to receive their rightful aid from these agencies, which in turn, we strongly believe, would improve their life at adulthood [4]. In case of a sudden outbreak staging an instantaneous survey to an outsized group of citizens now become realizable. Collection of data can now be done remotely, thus reducing the likelihood of disease spread among the concerning data collection authorities. The system briefly brings down a busy task of data collection to the citizen's fingertips.

The future scope includes the security aspects of the web site and improvement of the machine learning algorithm for disease prediction. Also, a cloud service might be connected to the system, to collect together all the incentives and schemes found out by the government or NGOs for the people.

6. References
[1] S. Vyas, M. Gupta and R. Yadav 2019 Amity International Conference on Artificial Intelligence (AICAI) pp. 709-711
[2] D. Dahiwade, G. Patle and E. Meshram, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) pp. 1211-1215
[3] B. Narendra Kumar Rao, B. Bhaskar Kumar Rao, Vellingiri J 2019 International Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume-8 Issue-8 June ISSN 2278-3075.
[4] Abhishek Verma, Sunita Aggarwal, Sandeep Garg and Pooja Anand 2019 Journal, Indian Academy of Clinical Medicine Vol. 20, No. 3-4
[5] Angell BJ, Prinja S, Gupta A, Jha V, Jan S 2019 The Ayushman Bharat Pradhan Mantri Jan Arogya Yojana and the path to universal health coverage in India: Overcoming the challenges of stewardship and governance (PLoS Med16(3): e1002759).
[6] J. Li, P. Ray, H. Seale And R. Macintyre, 2012 45th Hawaii International Conference on System Sciences (MAUI, HI) PP. 2800-2809
[7] J. Li and P. Ray, 2010. The 12th IEEE International Conference on e-Health Networking, Applications and Services pp. 391-398.
[8] Dandona R, Pandey A, Dandona L. 2016 Bull World Health Organ. 94(4); 286-96A.
[9] Devadasan N, Ghosh S, Nandraj S, Sundararaman T 2014 “Monitoring and Evaluating Progress towards Universal Health Coverage in India.” PLoS Med 11(9); e1001697.
[10] Jones M, Johnson M, Shervey M, Dudley JT, Zimmerman N 2019 J Med Internet Res;21(8): e13600
[11] https://www.kaleido.io/blockchain-blog/what-are-smart-contractsand-how-do-theywork/ “What are Smart Contracts and How Do They Work?”
[12] https://solidity.readthedocs.io/en/v0.6.8/, “Solidity Documentation”
[13] https://www.trufflesuite.com/docs, “Truffle Suite Documentation”
[14] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534.