Section A-Research paper
ISSN 2063-5346

Sharing Personal Health Data by Blockchain with Cloud Storage Technologies

Balamuralikrishnan G[1] Assistant Professor, Dept of CSE, VCET
Balakrishnan G[2]Associate Professor, Dept of CSE, FMCET
V.Rajeshkannan[3] Assistant Professor , Dept of CSE, KCET
DR R.Ramya[4],Associate Professor, Dept of CSE, KCET
A.Alagar[5] ,Assistant professor ,Dept of CSE, SIT

Abstract—The development of wearable technology and mobile computing leads to the generation of huge amounts of personal data. Every second, many data related to health is being created and amassed continuously. As these personal datasets include important information about the users, they have to be treated as an asset for these users. Also they have to be controlled and managed by the person who generated those data. Many service providers have been established to manage these datasets and to keep them safe. These service providers project the problems for data security and prevent data exchange. These individual health records are very important sources. In this paper, we exhibit a theoretical approach for sharing continuously dynamic personal health data leveraging block chain technology and cloud storage to share health-related data in an open and secure way. In addition, to have control over data quality, we also develop a data quality inspection module based on machine learning approaches The proposed system's main objective is to give users the ability to possess, manage and share their personal health data in a secured way that go along with the General Data Protection Regulation (GDPR) additionally, it gives users of commercial data and researchers an effective approach to get high-quality personal health data for both of those things.

INTRODUCTION

People and technology have become inseparable as they were using various kinds of smart devices like smart phones, smart watches, smart bands. These devices use various health-related applications such as remote diagnosis. We have to be grateful to the rapid development of mobile computing , wearable technology and wireless sensing. This helps us to monitor diseases and to look after elderly person as it predicts the disease earlier and make people health conscious. These devices provide a significant amount of data on a person's health, and this data is useful for both academic and commercial healthcare research. proper sharing of personal health data by the patients, researcher's, business's and also by the entire public health care system is always important Health data should be owned and managed by the respective users themselves as a personal asset but in practise they are frequently managed by various service providers, device manufacturers, or dispersed across several healthcare systems. As these centralised data warehouses and authority providers are desirable targets for cyberattacks, it generally creates obstacles for data sharing and compromises data security and privacy. Due to cryptocurrencies, the blockchain technology has significantly increased in popularity in recent years, particularly in the financial sector. For instance, since its initial introduction in 2008. Bitcoin has drawn interest from the research community across a range of academic disciplines and has become widely accepted because of its distinctive features, such as lack of control which is centralized, an assumed high of innominateness and distributed consensus over decentralised networks. Multiple parties' cooperation is prerequisite for the decryption of the data. With that cooperation blockchain solutions could reduce the risks of data breaches. This is also possible by using threshold encryption of data combined with public key framework. The blockchain-based data sharing system might significantly streamline the process of acquiring data for research and commercial initiatives and give users the chance to acquire ownership and privileges over their own data and profit from them. Additionally, it provide greater data control to help the user to keep an eye on all the actions using those data. The purpose of this paper is to put forward the significance of sharing personal health data based on blockchain and cloud

storage technologies . this paper also highlights how these technologies provide safe and secured sharing system. it also brings out how these systems backup researchers and consumers of commercial data in obtaining the data they demand efficaciously in a clear manner and in union with data regulations like GDPR.

II. RELATED WORK

The way of using personal data created by mobile and wearable technology has always to develop the standard of health care system. generated immense interest among researchers. One of the barriers for these investigations is data collection as it is naturally costly and time consuming . the worries about the security of their data make most people to keep their medical and other data related to their health confidential and are not ready to share them, A decentralised network of peers and a public ledger can be used to provide reliable and auditable computing, as shown by the financial industry's adoption of blockchain technology. Recently, there have been a lot of research about using blockchain technology in industries other than finance. The research in from 2015 utilised blockchain to safeguard data privacy for individuals. A protocol developed by the authors changes a blockchain into an automatic access-control manager that ensures the users the complete control of the data and it eradicate the dependency of the users on third party for data sharing . the utilization of blockchain technology in manging healthcare data has been focused in many research papers since 2016.the research described in proposed an application structure based on blockchain called Healthcare Data Gateway (HGD). This provide a system which guarantees the patients to share their data securely and provide a complete control over the data sharing without involving privacy.it provided an efficient means of improving health care systems intelligence in the process of protecting the privacy of patient data. Electronic Health Records(EHRs) manages the data using blockchain technology. The study in [3].created a decentralized management system to maintain the health record called MedRec. This system enables easy access to the medical records of the patients across multiple doctors and various sites for treatment and also it provide a clear log of their medical history. The studies described above mostly focused on utilising blockchain to manage static health data like Electronic Medical Records (EHRs) (EMRs). The EMR comprises information that is mostly constant over the course of a patient's life, such gender, blood type, fingerprint, etc., or that changes gradually, like age, weight, height, disease history, etc.

The ability to store and share data inside the blockchain is made possible by the fact that this form of data often requires less storage space.This kind of static data only makes up a small portion of the total health data in the majority of real-world healthcare applications. Mobile and wearable devices, which are extensively used, generate significant volumes of dynamic data with frequent and big data sizes. For instance, the accelerometer within a smart watch produces data at a high frequency, producing millions of records every day with a maximum data size of several terabytes. It is challenging to store and share directly inside the blockchain because to the high changing frequency and size. A roadmap for a blockchain-enabled, decentralised personal health data ecosystem was put forth in a recent paper published in [1]. In order to help regulators overcome their issues and give people back control over their personal data, including medical records, the authors proposed the idea of a safe and transparent distributed personal data marketplace using blockchain and deep learning technologies. To offer an offchain storage option for huge biomedical data files, they integrated cloud storage into the ecosystem. In contrast to the majority of prior blockchain applications, a unique position called a data validator was added alongside the conventional roles of data contributor/generator and data consumer. The purpose of data validators is to verify or certify the accuracy of the data provided or generated by the users. Only data that has been verified or confirmed by data validators is made available to the public. The issue of gaining control over the data quality is resolved by this approach. Although most static and gradually changing health data may be controlled using this way, handling high frequency, large-volume data, like that from accelerometers, remains difficult. Manually validating these types of data may be time-consuming and require a lot of effort. Advanced techniques, such as machine learning and data mining, are required to manage such a massive volume of data. For many years, there has been a lot of interest in the research on evaluating massive data generated by wearable technology [4], [5], [6], [7], and [8]. For instance, the study of [9] assessed the degree of tremor in individuals with Essential

Tremor using the acceleration data obtained from a smart watch.

For Human Activity Recognition (HAR) tests, the researchers used a deep learning algorithm in [10] and found satisfactory results can be used to assess the calibre of data generated by wearable and mobile devices, similar to deep learning. In this paper, I presented a noval method for sharing personal health data that is enabled by blockchain, cloud computing, and machine learning technologies. This method was encouraged by the before mentioned studies.

III RESEARCH SCOPE

According to [1], health data can generally be split into dynamic and static data. The term "static data" refers to personal information that has remained largely constant over the course of a person's life, such as a person's DNA or fingerprint. The dynamic data shows the user's activities throughout time, such as electroencephalogram (EEG) heart rate data, or the status of the organism at the time of sampling, such as blood test data.

The dynamic data can be further broken down into variables that change quickly, like acceleration variables, and variables that change gradually, like height and weight variables, etc.

The health data can be separated into continuous data and instant data depending on the data gathering method. The continuous data is gathered over a period of time and shows the user's status or activities throughout that time. Continuous data are typically dynamic time series data that change quickly. The instantaneous data, in contrast, are acquired through a single measurement. The instant data may be dynamic data that changes over time or immutable static data. The classification methods are not based on the health indicators that the data represent, but rather on the type of data and the data gathering techniques. The same health indicator's data may fall under several categories. The number of heartbeats per minute, for instance, can be used to convey the status of the heartbeat and is an example of instantaneous dynamic data. However, when the heartbeat is monitored with EEG over a period of time, the data that is gathered is still of the same heartbeat but is continuous-dynamic

This study mainly focus on the continuous -dynamic data as it makes up the majority of the data created by wearable and mobile devices. The below mentioned figure picturizes this sort of data. There aren't many research on this subject because they typically occur frequently, are big in size and it is hard to be shared or saved using the similar techniques used to store other types of health data

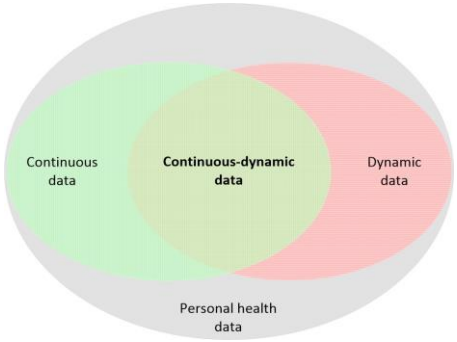

Fig. 1. Personal data categories and the study scope of this paper

A. General Data Protection Regulation:

The new General Data Protection Regulation (GDPR) [2] of the European Union went into force in May 2018. The present data protection directive, which was enabled in 1995 will be replaced by this GDPR. This is one of the most significant updates to data privacy regulation in recent memory. The regulation's main objective is to harmonise data privacy regulations across Europe, with a focus on empowering and safeguarding the privacy of EU individuals. The problem of user permission is one of the most important ones. The rule stipulates that the service provide should give a clear explanation of the purposes for the usage of the user’s consent and similarly the user must have the same ease in giving and withdrawing consent. If the user withdraw his consent, the service provider is supposed to destroy the data related to that particular user. In addition, the user has the right to access, which requires that, upon request, the service provider or business disclose whether or not it is processing the user's personal data as well as its intended use. All information must be made available to the user by the service provider in a machine-readable format. The user has the right to get the information from the controller in a machine readable format and can transfer it to another controller .this right to data portability is alike to the right of access. Companies which break the above mentioned rule will be subject to fines up to 20 million euros or 4% of their revenue.

CONCEPTUAL DESIGN

The proposed health data storage and sharing system's design has been depicted in the figure 2. In this system there are three defined roles

• Users: to create, upload, and sell personal health data in exchange for money or services.

• Key keepers: to store the secret keys for encrypting and decrypting data after a user uploads it, then release the keys to users once a transaction is authorised. Every transaction that is accepted will result in financial rewards for them.

• Customers: to purchase user data and offer users and key keepers financial or service incentives.

A related App was created for each role to assist users in achieving their goals.

Users: User app is a mobile computing service such as a smartphone or tablet. It can transfer with different health-related data, wearable devices or sensors such as Bluetooth. The corresponding sensors' APIs provide support for data collecting. The quality score of the raw data can be get by passing them through a quality validation module. A brief discussion about this module will be done later. After validation , the raw data will be combined with different identification marker.

To simply describe the data, use the title, Data Type is used to specify the kind of the data, Size is used to specify its size, Quality is used to specify its calibre, etc. If necessary, certain static personal information, such as gender, age, and weight, could also be incorporated. The next step is that the combined data will be compressed and encrypted. Then, the encrypted data will be put on a cloud. Data encryption keys will be divided up into shares and given to key keepers. Following that, a transaction will be created and disseminated to all blockchain nodes. The transaction includes the user's public key, a hash reference that connects to encrypted data, basic details, the dataset's price, etc

Key keeper App: This app generally runs on a local device that is connect to the internet service or on a cloud server. This app will keep the key shares securely which is received from the system. After validation of the dat shares, it will get a notice to release the related key share to the customer who made that transaction.

Customer App: this app is also similar to the keykeeper app as it connects to the internal or on a cloud server. This provide all the available datasets to the customer and permit them to close the type of datasets they prefer. After choosing the dataset, the customer will be guided by the app to make a transaction for buying that dataset. then this transaction is validated the customer app will get a link to the encrypted data and the decryption keys from the key keepers. Finally ,it enables the customer to download the date and decrypt them with the private keys.

The following sections describe the core modules and its function of the proposed system.

1.validation of data quality:

This study mainly focus on continuous - dynamic data. Typically, these data are produced by common sensors. The devices' APIs make it possible to access the sensor data.

To ensure that the data is valid in accordance with specific validation patterns or checks, advanced machine learning techniques can also be used to evaluate the pattern of the acquired data. We can use it to verify the data's accuracy from a hardware and software perspective.

Hardware aspect: once, a new device is linked to the user app, the user app will gather hardware details about that device and any embedded sensors. A device or sensor is recognised as certified hardware and the data it generates is dependable if it comes from a validated manufacturer. Otherwise, the App won't allow it to connect. A well-maintained database of approved manufacturers and gadgets should be created in advance for this purpose.

3.Software aspect: With the aid of cutting-edge machine learning methods, it is feasible to accurately categorise the patterns in a time series collection. Numerous studies have been conducted on this subject. For example, it can identify a user's daily activities using information gathered from an accelerometer built into wearable technology [9], [10], [11]. We can build reliable classifiers for various health data using comparable machine learning techniques. The nonsensical data and sounds will be removed, leaving only the data with predetermined features saved. Here, a relative criteria for data quality is used. Consider the aforementioned acceleration data as an example and visualise how a smart watch would collect the user's acceleration data over the course of a 24-hour

period. Sleep will be distinct from other daily activities in the quality validation methods. The data corresponding to the sleep time may be categorized as high quality data or noise based on the user's decision to share data related to sleep or daily activities ,

2) Data sharing transaction validation: The blockchain module is one of the system's main building blocks. It's employed to protect the data-sharing procedure. Ethereum1, Hyperledger Fabric2, and other decentralised blockchain application platforms are currently accessible. These platforms make it simple for developers to make blockchain applications. In this study, we decide to use Ethereum as the system's development framework. Developers can create and issue their own cryptocurrency or tradeable digital token on the Ethereum platform, which can be used as money, a proxy for assets, or a digital share. The contract will be instantly compatible with any wallet, other contract, or exchange that also uses this standard because these tokens use a standard coin API.

Another crucial factor in selling or exchanging personal data with data consumers or commercial entities is that the data must be adequately anonymized to comply with GDPR laws [2]. For instance, personal demographic information like name, address, and person identifiers will be correctly erased or hashed so that the final dataset supplied to the consumer complies with data standards like GDPR [2].

Cloud storage : the main objective of including cloud storage in the data sharing system is that it provides a huge dataset as an off-chain storing option. In a long-term process, continuous dynamic data are often acquired at high frequency. The following accelerating data can be considered as an illustration. The data gathered in a single day may be several terabytes.

4. Data encryption: To maintain security and privacy, the user App will encrypt the data before uploading it to the cloud using symmetric-key techniques like Rijndael AES in conjunction with a threshold encryption scheme [1], and other methods. Then, using Shamir's secret sharing approach , the symmetric key for decrypting the data will be divided into several shares, and the key shares will be given out to various key keepers. The total number of key keepers and the blockchain security model define the required minimum number of key keepers to decode the data [1]. In order to access the encrypted data, The link and authentication to the data must be obtained.Then, in order to decrypt the data, he or she must obtain sufficient key shares of the encryption key. Theoretically, individuals can only access these details through a transaction that has been validated and confirmed by blockchain nodes.

5. Crypto token: It is necessary to offer users incentives when they submit their personal data in order to motivate them to do so. Health-related services, such as disease monitoring and diagnosis, might be advantageous, but this approach might not be appropriate in every circumstance. In some circumstances, it is preferable to offer some sort of

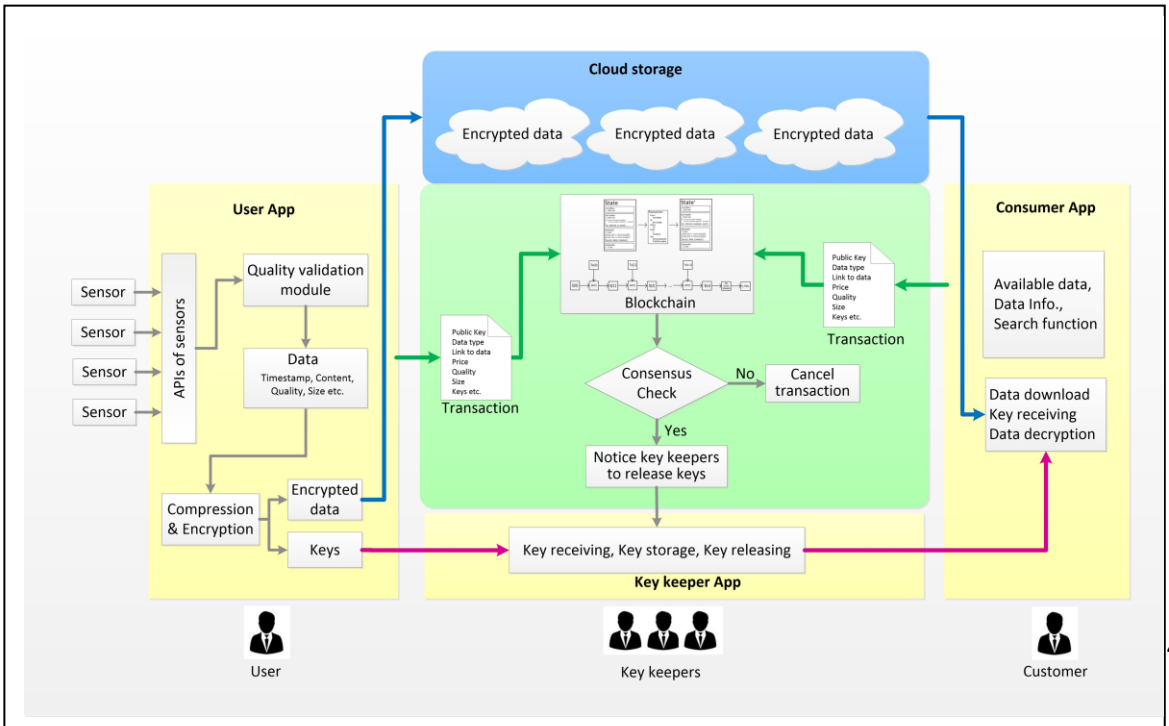

Fig. 2. The general architecture and workflow of the proposed system

financial benefit.

For a variety of reasons, such as the requirement to carry out an enormous number of microtransactions across numerous zones and among an equally enormous number of different types of participants, exchanging personal information for currency may be problematic. In line with [1], we recommend our own cryptocurrency token called Personal Health Data coin (PHD coin), which can be created or mined by storing data on a system with a blockchain to help with transactions. When the network has enough nodes and users, it is anticipated to facilitate exchange with other digital currencies or real currencies in the future.

6. Overall workflow: As seen in figure 2, there are three pipelines involved in interactions between users, key keepers, and customers: blockchain transactions, encrypted data transfers, and key sharing for decryption. The general process looks like this:

• Step 1: Using the user app, the user gathers, compresses, and encrypts data. The information will then be uploaded to a cloud storage system, and the key shares will be sent to the key holders.

• In step 2, a transaction will be created to inform the other users that the data has been uploaded and is ready for sharing. This transaction will be added to the blockchain via a consensus process after the user confirms it, and consumers and data verifiers will be able to see it.

• Step 3: The data validator will run data-mining and machine-learning algorithms on the data as part of the validation of the data to ensure that the data is valid in accordance with rules and specifications. The data validator will then certify that the data is valid in accordance with the provided guidelines.

• In step 4, a client selects the data they wish to purchase, confirms that it has been approved by the necessary data validators, and then establishes a transaction to do so.

The customer will sign the transaction, and it will then be added to the blockchain.

• In Step 5, the data purchase transaction will be verified using the consensus algorithm. The transaction will be allowed if the consumer has adequate balance (PHD coins) to pay for the data. A specific quantity of PHD coins will be given to the smart contract in accordance with the price of the data, and the workflow advances to step 6. Otherwise, the transaction will be declined, and step 3 of the procedure will be resumed.

The key keeper of the associated dataset will be notified in Step 6 that a transaction to buy the data has been authorised. Then, using an authenticated communication channel, key keepers give the client access to their key shares of the relevant dataset.

Step 7: A predetermined rule will be used to the distribution of the PHD currencies saved in the smart contract among the accounts of the user and the key keepers.

• In step 8, the customer is given access to the encrypted data through a link and sufficient key shares to decrypt it. The data can then be downloaded and decrypted by the user from the cloud storage. The workflow comes to an end since the data are usable.

DATA SECURITY ANALYSIS

The data security of the proposed system depends on the following setups

The data kept in the cloud provide limited access to the user. One has to know the address and get authentication in order to access the encrypted data. the data is encrypted before being uploaded to the cloud storage, inorder to check the storage system from data leakage single symmetric key is used by multiple key keepers for the decryption of data. As a result, the compromise of the data.

The data transaction process is further secured by the use of the hash function and public-key signature techniques in blockchain contracts. without the permission of the user no personal information will be involved as it provide the facility to use pseudonyms.

The blockchain-based public key infrastructure(PKI) enables key keepers to generate verified communication channels with another participants. It secures the symmetric decryption key for the transmission and storage of data

The procedure of sharing data was the main subject of this investigation. The main focus of this work is not the data security concerns that arise after the data has been purchased and delivered to the customer, such as data leakage caused intentionally or unintentionally by the customer. Existing security mechanisms for data in use and at rest [11] could be used to provide this protection, but it is outside the focus of the current paper

CONCLUSION

In this research paper,I presented a blockchain-based, cloud-based, and machine learning-based method for sharing personal health data. It enables consumers to easily and securely own, control, and share their personal health data while also gaining advantages. Within the framework of health-related data from wearables and mobile devices, we initially divided personal health data into various groups based on data characters (dynamic and static data) and the data gathering methods (continuous and instant data). We suggested employing various approaches to transfer the continuous dynamic data of huge size using hash references to the storage place.

Second, by combining blockchain and cloud storage, our suggested approach got over the continuous-dynamic health data's size restriction. Additionally, I suggested that huge amount of health-related data may be saved on cloud and shared as metadata or transactional data. A data quality validation module was also included in this system as it control the data quality from the aspects of both hardware and software supported by machine learning techniques.

REFERENCES

[1] P. Mamoshina, L. Ojomoko, Y. Yanovich, A. Ostrovski, A. Botezatu, P. Prikhodko, E. Izumchenko, A. Aliper, K. Romantsov, A. Zhebrak et al., "Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare," Oncotarget, vol. 9, no. 5, p. 5665, 2018. 1, 2, 5, 6

[2] The European Parliament, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46 (general data protection regulation) [GDPR]," Official Journal of the European Union, vol. 59, no. L119, pp. 1–88, 05 2016. 1, 3, 4

[3] A. Ekblaw, A. Azaria, J. D. Halamka, and A. Lippman, "A case study for blockchain in healthcare:"medrec" prototype for electronic health records and medical research data," in Proceedings of IEEE open & big data conference, vol. 13, 2016, p. 13. 2

[4] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, "Elderly activities recognition and classification for applications in assisted living," Expert Systems with Applications, vol. 40, no. 5, pp. 1662–1674, 2013. 2

[5] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," Expert Systems with Applications, vol. 59, pp. 235–244, 2016. 2

[6] C. Pulliam, S. Eichenseer, C. Goetz, O. Waln, C. Hunter, J. Jankovic, D. Vaillancourt, J. Giuffrida, and D. Heldman, "Continuous in-home monitoring of essential tremor," Parkinsonism & related disorders, vol. 20, no. 1, pp. 37–40, 2014. 2

[7] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep activity recognition models with triaxial accelerometers." in AAAI Workshop: Artificial Intelligence Applied to Assistive Technologies and Smart Environments, 2016. 2

[8] M. Ermes, J. Parkk ¨ a, J. M ¨ antyj ¨ arvi, and I. Korhonen, "Detection of ¨ daily activities and sports with wearable sensors in controlled and uncontrolled conditions," IEEE transactions on information technology in biomedicine, vol. 12, no. 1, pp. 20–26, 2008. 2

[9] X. Zheng and J. Ordieres-Mere, "Detection and analysis of tremor ´ using a system based on smart device and nosql database," in Industrial Engineering and Systems Management (IESM), 2015 International Conference on. IEEE, 2015, pp. 242–248. 2, 4

[10] Y. Chen and Y. Xue, "A deep learning approach to human activity recognition based on single accelerometer," in Systems, man, and cybernetics (smc), 2015 ieee international conference on. IEEE, 2015, pp. 1488– 1492. 2, 4

[11] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on. IEEE, 2014, pp. 197–205. 4

[12] J. Daemen and V. Rijmen, The design of Rijndael: AES-the advanced encryption standard. Springer Science & Business Media, 2013. 5