2022

A privacy preserving online learning framework for medical diagnosis applications

Trang Pham Ngoc Nguyen
Edith Cowan University

Edith Cowan University

Copyright Warning

EDITH COWAN UNIVERSITY

A privacy preserving online learning framework for medical diagnosis applications

Author:

Trang Pham

NgocNGUYEN

Supervisor:

Iftekhar AHMAD

Daryoush HABIBI

Viet Q. PHUNG

This thesis is submitted for the degree of

Master of Engineering Science

in the

School of Engineering

January 10, 2022

Declaration of Authorship

I, Trang Pham NgocNGUYEN, certify that this thesis does not, to the best of my knowledge:

- Any content already submitted for a degree or diploma at any institution of higher education is incorporated without acknowledgement.
- Contain any previously published or written information by another person except where due reference is made in the text; or
- Comprise any defamation material.

I also give the Library at Edith Cowan University permission to create duplicate copies of my thesis as needed.

Signed: [redacted]

Date: 23.08.2021

Abstract

Electronic Health records are an important part of a digital healthcare system. Due to their significance, electronic health records have become a major target for hackers, and hospitals/clinics prefer to keep the records at local sites protected by adequate security measures. This introduces challenges in sharing health records. Sharing health records however, is critical in building an accurate online diagnosis framework. Most local sites have small data sets, and machine learning models developed locally based on small data sets, do not have knowledge about other data sets and learning models used at other sites.

The work in this thesis utilizes the framework of coordinating the blockchain technology and online training mechanism in order to address the concerns of privacy and security in a methodical manner. Specifically, it integrates online learning with a permissioned blockchain network, using transaction metadata to broadcast a part of models while keeping patient health information private. This framework can treat different types of machine learning models using the same distributed dataset. The study also outlines the advantages and drawbacks of using blockchain technology to tackle the privacy-preserving predictive modeling problem and to improve interoperability amongst institutions. This study implements the proposed solutions for skin cancer diagnosis as a representative case and shows promising results in preserving security and providing high detection accuracy. The experimentation was done on ISIC dataset, and the results were 98.57, 99.13, 99.17 and 97,18 in terms of precision, accuracy, F1-score and recall, respectively.

Keywords: Blockchain, Private blockchain, hyperledger fabric, healthcare, deep learning, convolutional neural network, data privacy, security

Acknowledgements

First and foremost, I would like to express my heartfelt appreciation to Dr Iftekhar Ahmad, my principal supervisor, for his continuous support, his patience and the constant motivation he has given me in this project.

I also would like to especially thank to Professor Daryoush Habibi, my co-supervisor. I really appreciate his insightful opinions, observations, and valuable suggestion on addressing the key points of the research works.

My special thanks to Dr Quoc Viet Phung for his technical contributions. During my journey, he has always supported and guided me, sharing his knowledge, and helping me learn and become a better researcher over the duration of my Master's program.

Last but not least, I am sincerely grateful to my family; my husband, my parents, and my children for their unconditional support, patience, sacrifice and continuous encouragement during the master journey. Without them, this achievement would not have been possible. Thank you.

Contents

Chapter 3 is not available in this version of the thesis

List of Figures

List of Tables

List of Abbreviations

PoW	**P**roof **o**f **W**ork
EHR	**E**lectronic **H**ealth **R**ecords
PII	**P**ersonal **I**dentity **I**nformation
PHI	**P**ersonal **H**ealth **I**nformation
CNN	**C**onvolutional **N**eural **N**etwork
NN	**N**eural **N**etwork
ISIC	**I**nternational **S**kin **I**maging **C**ollaboration
CC	**C**hain **C**ode
SC	**S**mart **C**ontract
CA	**C**ertificate **A**uthority
AI	**A**rtificial **I**ntelligence
ANN	**A**rtificial **N**eural **N**etwork
SVM	**S**upport **V**ector **M**achine
SMC	**S**ecure **M**ulti-party **C**omputing
DCNN	**D**eep **C**onvolutional **N**eural **N**etwork
ILSVRC	**I**mageNet **L**arge **S**cale **V**isual **R**ecognition **C**hallenge
RGB	**R**ed **G**reen **B**lue

List of Publications Arising From This Thesis

- T. Nguyen, I. Ahmad, D. Habibi and Q. V. Phung, "Skin Cancer Classification Using Transfer learning," under review, The Journal of Healthcare Informatics Research, Springer, 2021.
- T. Nguyen, I. Ahmad, D. Habibi and Q. V. Phung, "Privacy-preserving online learning using private blockchain" manuscript in preparation for the Procedia Computer Science, Elsevier Science, 2021.

This thesis is dedicated to my parents.

Chapter 1

Introduction

The advancement of Information and Communication Technology (ICT) has ushered in a new era in which information plays the pivotal role in almost all aspects of our daily lives. Agriculture, industrial automation, smart cities, and healthcare, amongst other areas, rely heavily on new technologies and communication systems. Healthcare is the most essential application amongst these, as it is one of the most basic human needs.

In the healthcare sector, medical records and information are often scattered across multiple departments and systems. The requirement for accurate and valuable information is increasing, where patients need to be confident that their medical data is being stored securely. Healthcare data security plays a crucial role in protecting privacy and preventing data compromises [1]. Medical information is progressively and predictably converted into standardized formats in the form of EHRs [2]. To date, electronic health records have been gathered and managed by different authorized healthcare providers, often containing medical history, clinical data and pharmacy data. In today's healthcare systems, electronic medical records give each patient control of his or her medical data. Consequently, data can be employed in preprocessing methods for enhancing study environments [3].

So in this era, medical information is always an useful asset for patients as well as valuable resources for researchers and medical institutions doing disease study or conducting commercial transactions. Because medical information is a valuable asset, so patients want their medical information to be kept privately due to individual privacy concerns, however under specific circumstances, they may be willing to share partly private data to medical organizations. Even so, they are worried about medical privacy throughout the data sharing process.

To solve this problem, in this thesis, a blockchain-based privacy-preserving online diagnostic framework is presented. This framework combines a private blockchain with a machine learning model that helps to classify an important medical condition and still secure the data privacy for all medical information.

1.1 Background

Processing huge amounts of health data was a major problem in the early days of the digital era, leading to the adaption of machine learning in the biology domain. Since then, biomedical and biology sciences have progressed to new levels by identifying relations and exploring new information that was not considered previously. Accordingly, machine learning has become popular across the healthcare industry, from extracting information from medical data to predicting or diagnosing disease. The incorporation of machine learning techniques into the computational biology field has considerably enhanced medical diagnosis techniques. Many illness diagnoses are currently made utilizing medical image processing techniques and machine learning algorithms. Additionally, machine learning based computational decision

making has become common in resource allocation, patient care, and research on treatments for various illnesses.

To date, machine learning algorithms have assisted in data analysis in a wide range of disciplines. These new and sophisticated approaches can change medical treatment by utilising well-trained classification models to accurately and quickly detect, prevent and treat illnesses [4, 5]. Developing a well-trained model often requires a large volume of sensitive health information during the training phase. Collecting a vast volume of medical data and training a complex diagnosis model, can be challenging for a single hospital. For example, owing to geographic constraints imposed by some uncommon diseases, a single healthcare facility may lack sufficient patient health information data set, resulting in an unbalanced training dataset that reduces the classifier's performance. One effective technique for developing a disease diagnostic learning model is to gather medical records from multiple hospitals, store them on a centralized server, and then train a classification model in the traditional manner.[6].

Figure 1.1 has been removed from this version of the Thesis. The figure can be viewed at
https://www.researchgate.net/figure/Centralized-Multi-party-Learning-Master-Worker-Pattern-vs-Decentralized-Multi-party_fig1_330625714

FIGURE 1.1: Centralized learning (Master-worker Pattern) vs. Decentralized learning

Transferring such a vast amount of data to a central server, may result in massive communication overheads as well as serious security and data

privacy issues. Federated learning with sufficient security measures can potentially offer a solution in this context. Federated learning [7], often referred to as distributed learning or collaborative learning, is a hybrid of distributed computing and deep learning in which a parameter server maintains a deep learning model for training, and several participants engage in the distributed training process. A basic federated learning paradigm is a master-worker pattern as shown Figure 1.1 (centralized). In the paradigm, the participants separately calculate changes to the existing model using their local dataset and submit them to a computing server, which then coordinates the learning process and combines the results into a new global model. The effectiveness of both these techniques is highly reliant on the central server's integrity, which is unfortunately vulnerable to attackers and, consequently, threatens the privacy of the training data. Worse still, the centralized server itself might be malevolent, attempting to steal critical information from the training data for financial gain.

Blockchain [8] has been introduced as an innovative and secured approach to store information, execute transactions and build trust in digital environments where every member is treated equally. Blockchain technology is a decentralized ledger that generates distributable and immutable records that are exchanged between network database systems. Accordingly, it is a network of decentralized nodes (Figure 1.1 - Decentralized), where the information of each node is encrypted with the code of its ancestor node [9]. Due to the fact that it is an immutable and shared ledger, the data contained inside becomes accessible to everyone. As a result, anything produced on the blockchain is visible and transparent, where all engaged parties are responsible for their activities.

With all these advantages, blockchain appears to be a viable answer to aforementioned untrustworthy central server issues and privacy problems.

Because of its inherent features, blockchain nodes can interact inside a completely decentralized network in which each node communicates in a peer-to-peer node and works asynchronously. Even though the training data is split and stored separately, federated learning architecture cannot ensure the data privacy of training data.

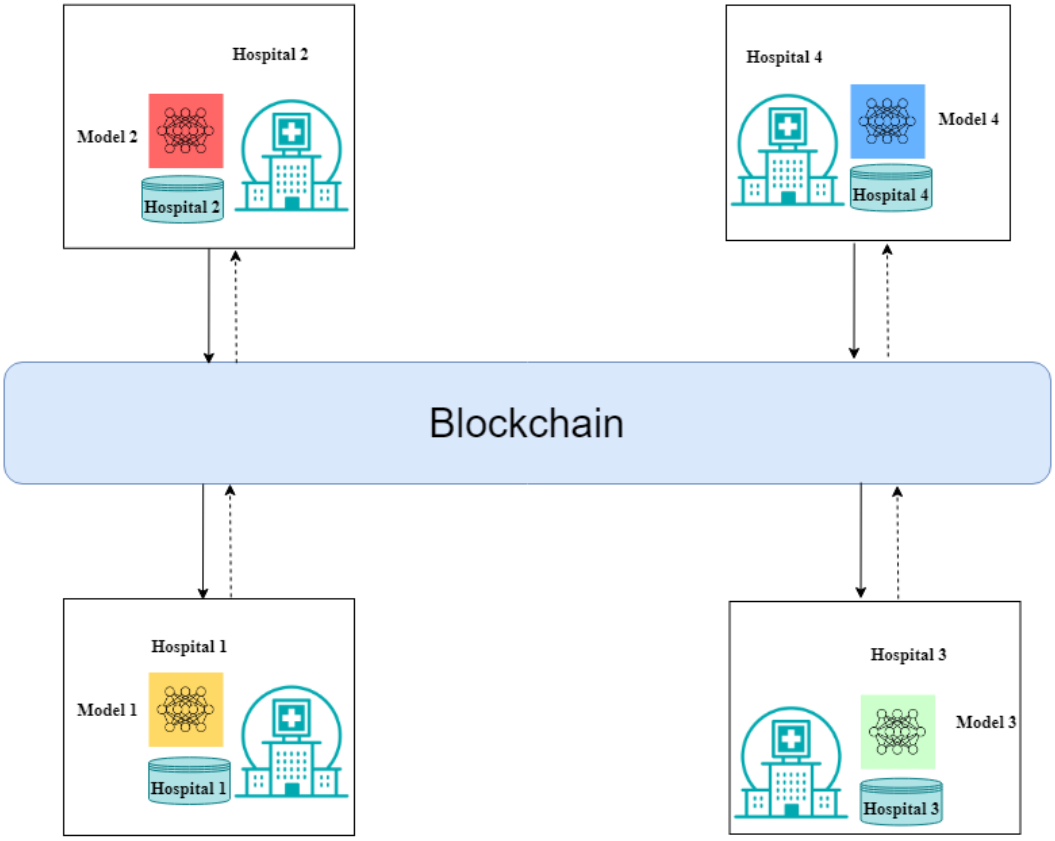

FIGURE 1.2: The Architecture of the system

To solve this privacy issue, and motivated by the revolution of blockchain in healthcare, **this thesis investigates and presents privacy protection for online machine learning framework that combines two important technologies, which, provides a provocative environment for healthcare institutions to collaborate and train an image classification model for skin cancer diagnosis**. Figure 1.2 presents the architecture of our proposed system based on blockchain, in which hospitals in the network have their own dataset, and their goal is to train a disease diagnosis model cooperatively.

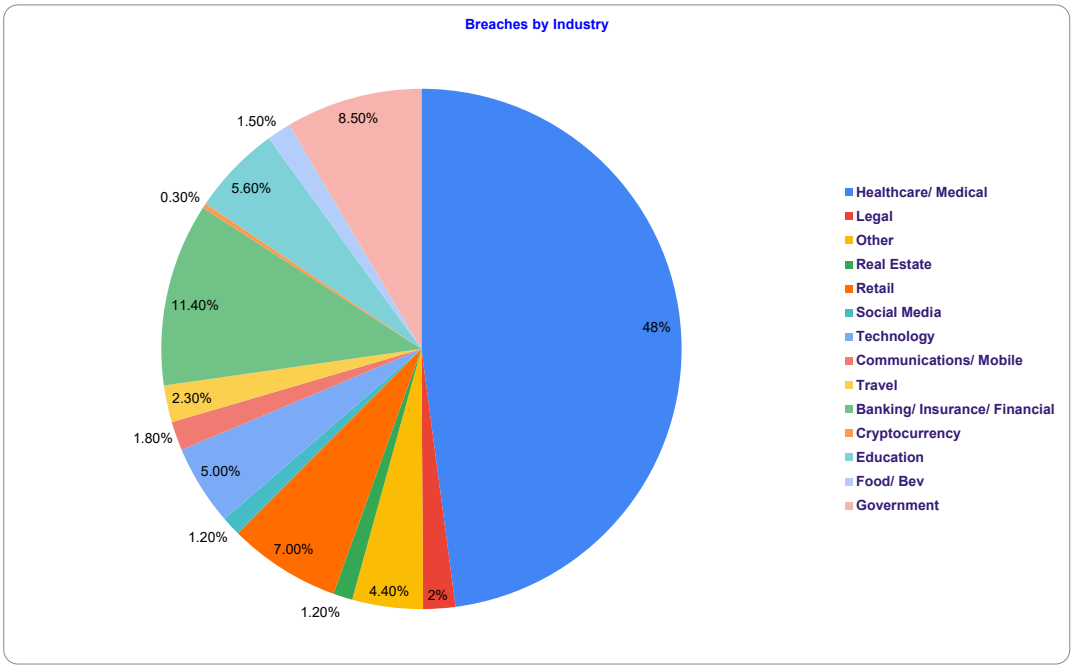

FIGURE 1.3: Data breaches by industry

1.2 Significance and motivation

Health, government, and financial information are the most commonly stolen data forms. As shown in Figure 1.3, in 2018 the health-care sector experienced more than four times as many breaches as any other sector, equivalent to almost half (48%) of all breaches [10]. In the healthcare industry, medical data sharing applications are one of the most alluring targets for cybercriminals, where data breaches were estimated at over 112 million records in 2015 alone [11]. Digital healthcare is becoming the new normal, with the telemedicine industry alone projected to cost about $41.2 billion by 2021.

Health data is valuable, so it faces an even larger challenge. According to a recent investigation [12], the quantity of medical records exposed each year has been increasing. Patient data not only contains a variety of personal identity information (PII) that cyber criminals can use to commit identity

theft, such as patient names and addresses, insurance, driver's license numbers, and financial account information, but it also contains extremely personal and private information about a patient's physical conditions, medical ailments, smoking status, disabilities, and mental health concerns.

Healthcare service providers have access to a massive database that functions as a repository of consumer information, one that is more expensive than in other industries or organizations. Thus, personal health information features as an attractive commodities for hackers and cyber-criminals. A data breach may be costly for most industries, but data health breaches can be fatal for relevant industries. Accordingly, data privacy for health information has emerged as an important task that needs to be solved.

By taking advantage of multiple predictive modelings in different cross-institutional healthcare sites, we could advance the research and stimulate the environment for quality improvement proposals, which is crucial for achieving the national healthcare department's targets. There are many privacy-protecting methods that have been carried out on the predictive models; however, the core mechanism is still based on a centralized architecture, which is susceptible to the security and robustness risks such as malicious record injection or single-point-of-failure. Online training, which allows hospitals or clinics to build shared prediction models collaboratively while keeping training data private, is critical in addressing the data privacy issue.

In this thesis, we present a new framework for integrating the blockchain technology with machine learning model for the purpose of privacy-protection. This system will use the features of blockchain technology, a distributed architecture that provides anonymity, security, integrity, longevity, immutability, and transparency. This thesis propose a proof-of-concept implementation of a system based on the Hyperledger Faric platform.

1.3 Research Objectives

The primary goal of this thesis is to devise a framework for secured and online training of machine learning models for distributed dataset while preserving the privacy of sensitive data. In particular, the framework is intended for a healthcare application where the patient health records are independently stored in the database at clinics, or hospitals. The machine learning model would be deployed and locally trained at the premises. The online training framework would have then provided a mechanism to exchange information to improve the performance of the models without compromising the data privacy. In order to achieve the objectives, the following sub-objectives were pursued:

1. Develop a novel machine learning technique to support online training and improve the performance of the model with limited datasets. The image classification for skin cancer will be used as a case study.

2. To create a secured online learning framework for privacy-preserving machine learning leveraging private the blockchain technology.

1.4 Thesis contribution

The following items summarize the contributions of this thesis:

- This thesis introduces an online machine learning platform that allows every site in the network to exchange a partial model without exposing patient information.

- A machine learning model for image classification using transfer learning to tackle the limited data issues is presented.
- A privacy-preserving online training solution is presented. This thesis presents how blockchain can be utilized to facilitate health data sharing without sacrificing privacy and security.

1.5 Organization of Thesis

The remaning chapters of this thesis are organized as follows:

Chapter 2 provides a literature review of blockchain technology, deep convolutional neural networks (CNNs) and other different existing works and techniques that have been implemented for similar problems.

Chapter 3 focuses on describing the classification of skin images based on pre-trained CNN architecture and transfer learning.

Chapter 4 explains the blockchain system architecture and provides detailed information about the state and transaction flow in the proposed system.

Chapter 5 summarizes the thesis conclusion and discussion of possible future work.

Chapter 2

Literature Review

2.1 Background

As discussed in the previous chapter, this thesis leverages the blockchain technology to create a privacy-preserving online machine learning framework, where a pre-trained CNN model is utilized to detect skin lesion images on datasets of all hospitals in the network. In this section, we present detailed information and background review on two key technologies: blockchain and machine learning.

2.1.1 Federated learning

In healthcare, there are numerous diseases that need to be recognised in their early stages in order to begin appropriate treatments. If not, these diseases may become incurable and fatal. As a result, complicated medical reports, medical data and medical pictures must be accurately analyzed at early stages, but with better accuracy. Humans may not be able to detect some abnormalities in some cases. Machine learning techniques are being utilized in healthcare for computational decision making in instances where

critical data analysis on medical data is required in order to discover hidden correlations or abnormalities that are not obvious to humans. Implementing algorithms to accomplish such tasks is challenging.

To obtain a well-trained learning model, a huge amount of health data is required. However, it is challenging for a single hospital to gather a huge amount of annotated samples and train a complex disease diagnostic model. A central server that stores health data from multiple healthcare institutions has been proposed. However, security and privacy issues have been raised as an important concern for a central server. To address these obstacles, federated learning might be a viable option to increase training efficiency and facilitate knowledge sharing without compromising data privacy.

Federated learning [7] is a method which has recently gained prominence due to its potential for learning from fragmented sensitive data. Instead of collecting data from several sources, it allows for a shared training model using a central server while maintaining the data in the original institutions. Federated learning accesses the current model and then updates it locally using the device's data. These trained models locally are then transmitted back to the central server from the devices, where they are consolidated, and a single enhanced and integrated global model is then transmitted back to the devices.

In a broader sense, federated learning enables machine learning techniques to gather experience from a diverse set of data distributed across several sites. The method allows different institutions to participate in model building without requiring them to disclose sensitive data directly. Throughout multiple cycles of training, the shared models can be exposed to a much broader range of data than any single institution has at hand. Federated

learning, in contrast, decentralizes machine learning by eliminating the requirement to pool data in a single place. Rather, the model can be trained in numerous iteration across multiple places.

Machine learning

Machine learning (ML) is the process of educating a computer system to make correct predictions in response to input data. ML is concerned with the creation of computer programmes capable of accessing data and using it to self-learn.

Machine learning is a subfield of artificial intelligence (AI) that aims to develop new algorithms for making predictions based on given data. ML uses training data to create general models that can detect the existence or absence of a pattern in test data. Machine learning algorithms cone in a variety of forms, including supervised, unsupervised and reinforcement learning. These categories are summarized below:

- **Supervised learning:** Supervised learning is a machine learning technique that involves training models using labelled data. In supervised learning, models must discover a mapping function that connects the input variable X to the output variable Y.
- **Unsupervised learning:** Unsupervised learning differs from supervised learning in that the datasets are not classified. The ML algorithm should analyze the similarities between object pairs in order to create a structure from unlabeled data.
- **Semi-supervised learning:** It is a subset of machine learning in which models are trained using a combination of sparsely labeled data and a

huge amount of unlabeled data. This method is a hybrid of unsupervised and supervised learning.

Machine learning is a vast area of research that overlaps and incorporates concepts from a variety of closely related areas, including artificial intelligence. This involves several forms of learning that are beneficial in several fields, including finance, the banking system, and healthcare.

Healthcare is one area where machine learning may have far-reaching social implications. In a rapidly developing market of fit bits, smart watches, and other gadgets that continuously collect and analyse a wealth of medical data, the use of machine learning to evaluate this data is gaining traction. Machine learning can be used in a variety of health-related applications; some of these applications include assisting doctors in developing more personalized prescriptions and treatments for patients, as well as assisting patients in determining when and if they should arrange follow-up appointments.

Machine learning has been shown to be effective in identifying patients who are more prone to recurrent diseases and for assisting in diagnosis. Machine learning can be used to assist in diagnosing and directing patients to appropriate therapy while keeping them out of costly, time-consuming emergency care facilities.

In healthcare diagnostics, machine learning can be used to do classification tasks on binary or many dependant variables. Several machine learning methods are mentioned below that have been shown to provide accurate diagnoses in healthcare:

- **Neural networks**

  The biological neural network is a crucial component of the human

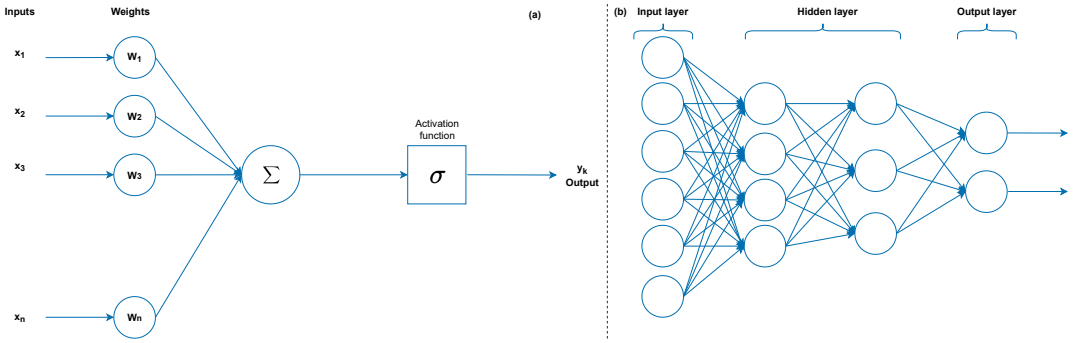

FIGURE 2.1: (a) is the perceptron layer, (b) is the image of multi-layer neural network

brain; it is a complex structure that can perform several functions at the same time. A neural network (NN) is a classification system that mimics the human brain and its neurons. Perceptrons are used as the fundamental unit of a NN instead of neurons (Figure 2.1 a). As shown in Figure 2.1 b, the NN architecture is made up of three layers: (1) the input layer, which contains input feature vectors; (2) the output layer, which contains the neural network response; and (3) the layer between the input and output layers, which contains neurons.

After training, a neural network can learn approximate target outputs by adjusting the weights of all neurons. Analytically solving neuron weights in a multi-layer network, however can be difficult. The back-propagation algorithm can be used to solve the weights iteratively in a simple and efficient manner. This algorithm calculates a gradient that is needed in the weights calculation.

Propagation and weight updates are the two steps of the back propagation technology. In the initial stage of this algorithm, an input vector is transmitted forwards through the neural network, whereby the output value is generated. Following that, the cost (in term of error) can be determined. Error values are then passed back to the network, which calculates the hidden layer neuron's cost. The neuron weights are then

modified in the second step of the algorithm by measuring the gradient of weights and subtracting the gradient of weights ratio from the current weight. The learning rate is the name given to this proportion. The algorithm continues with various inputs after the weights have been modified until the weights have converged.

- **Convolutional Neural Network**

  The convolutional neural network (CNN) is one of the most extensively used deep learning algorithms in the field of computer vision, where they are particularly good at image recognition since they can learn directly from images. This type of neural networks is comparable to the previously mentioned neural networks in that it contains neurons, weights, and biases. Each neuron receives an input and responds with an output.

  CNNs have one or more completely connected layers, similar to a multilayer neural networks, but they are simpler to train than fully connected networks with the same number of hidden neurons. CNNs can be employed to extract features from an input picture and use them to train a classifier since they can have hundreds of layers that each learn to recognize different features of an image. Each layer's output is then fed into the next layer [13].

  A CNN is composed of input and output layers in addition to several hidden layers. Typically, these hidden layers are composed of convolutional, pooling, and completely connected layers [14]:

  - Convolutional Layers: the results of the input are passed to the next layer by these layers. This mimics a neuron's response to visual stimuli.

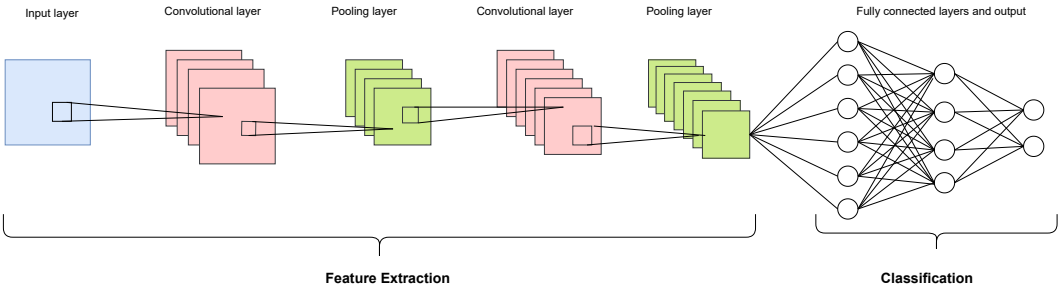

FIGURE 2.2: Example of CNN architecture

  - Pooling Layers: these layers integrate the outputs of the previous layer's neuron clusters into a single neuron in the following layer. This layer's objective is to reduce network parameters and computations.
  - Fully-connected Layers: these layers connect each neuron in one layer to every neuron in next layer.

- **Transfer learning**

  Deep learning, also referred to as deep structured learning, is a subset of machine learning that relies on a large number of algorithms. Since most current deep learning models are built on a NN, deep learning has a cascade of multiple layers similar to NNs.

  Deep learning is distinguished from traditional machine learning techniques by its ability Deep learning differs from traditional machine learning techniques in that it can derive meaningful features directly from pictures, sound, and text in unsupervised and/or supervised ways. Indeed, this method considers feature extraction to be a component of the learning process. With these deep learning features, the demand for hand-tuned machine learning methods decreases.

  Currently, the majority of deep learning applications, especially those involving computer vision, rely on transfer learning. Transfer learning is a type of machine learning algorithm in which a model that has been

trained on one job is reused on a related task that is connected. In the majority of cases in the medical area of computer vision, such as skin cancer diagnosis, a dataset is insufficiently large (e.g., there are only a few thousand images; however, CNN require many more), and training a CNN from scratch requires a significant amount of time. As a result, it is normal to train a network on a huge dataset (e.g., ImageNet in 1.2 million images) and then utilize that knowledge to initialize the network for the task at hand.

Transfer learning is a technique for determining when and how to employ the layers of a pre-trained model. This approach has been shown to be effective in a number of computer vision applications, even when weights are employed from completely unrelated domains.

There are two primary methods to apply transfer as follow:

- **Fixed Feature Extractor:** As a feature extraction mechanism, the pre-trained model may be used. This operates by omitting the final completely connected layer or the output layer and using the remaining network as a fixed feature extractor for the given dataset.
- **Fine-tuning:** Fine tuning is the process of making minor changes to further improve results. For instance, if there is a single dataset, it can be divided randomly into training and validation (testing) datasets using a ratio of choice. The model file can then be trained on the training dataset and subsequently on the testing dataset.

Blockchain

While machine learning can be a useful tool in healthcare, privacy and data security are a major concern. Innovative tools, like blockchain, can be useful in this regard. This section provides background information of blockchains.

A blockchain system serves as a distributed database where all participants are able to read and write data without exposing their own data [15]. It is comprised of a list of records that is constantly growing, referred to as blocks, which contain transactions that are recorded and added to the consecutive order [16]. Each node is linked to the network, and receives an automatically downloaded copy of the blockchain. These blocks are safeguarded against manipulation by cryptographic hashes and a consensus process [16]. Once data is saved on the blockchain, it cannot be altered. The blockchain is a public ledger that contains a verified record of every transaction that has occurred thus far [15]. Accordingly, it is a distributed database that eliminates the need for third-party valida tion and eliminates the need for a central authority. This section discusses the fundamental components of blockchain technology.

This will individually explore as structure, blocks, and consensus.

1. **Structure**: A blockchain is composed of successive blocks that may include a variety of transaction types. Each block is connected to its predecessors by its hash. Each hash uniquey identifies its corresponding block. Blockchains create block hashes using the SHA256 hashing algorithm.

2. **Blocks**: The longest chain, known as the genesis block, traces back to the very first block in a chain's history [17]. T o alter the sequence, an adversary must recalculate all blocks tracing back to the genesis block. This process involves a huge amount of computations, especially as the longest chain is always trusted [17].

3. **Consensus**: Consensus is a term that refers to widespread agreement. Consensus is the process by which ledger transactions are synchronised

and updated across the network in order for ledgers to change them only when transaction are validated by the correct participants; when ledgers modify them, they update and modify with the same transactions in the same order [16].

Consensus is a fundamental issue in decentralized systems. Requiring two or more agents to agree on a given value jointly. At the moment, certain agents cannot be trusted. As a result, the consensus process should be conditional. Different consensus methods, such as proof-of-work (PoW), proof-of-stake (PoS) and others , may be used with blockchains [18].

Hyperledger Fabric:

Hyperledger Fabric is a Hyperledger blockchain project lead by the Linux foundation [19]. The projects are all collaborative and open source, where they aim to create a cross-platform blockchain solution. Hyperledger Fabric aims to create distributed ledger solutions through the use of plugable components. Intended for corporate use, this blockchain allows for nondeterministic use cases, in contrast to permissionless blockchain with their limitations on cryptocurrency. Accordingly, general-purpose, distributed applications can be run without the need for cryptocurrentcy. The different components and properties presented in the following subsections have been gathered via Hyperledger Fabric documentation and articles describing this project in detail.

1. **Organization**: An organization defines peers in the network that operate in trust, much like a real-world organizations, consisting of multiple members.

2. **Chaincode**: also described as a smart contract is a program that handles business logic in a network. A chaincode is used to work with a ledger; enabling peers to create transactions that are run in a secured and isolated docker container.

3. **Peer:** A fabric network consists of peers, where each are given an identity by a membership service provider. The peers can select one of three roles: client/Submitting-client, peer, ordering-service-node/orderer.

4. **Channel:** Blockchain is a network of peer-to-peer communication paths between peers in a blockchain network. Transactions go through consensus with only the members of a channel, where they are not visible by any other peer outside of it. Peers who are not authenticated will be unable to access any information propagating across this channel.

5. **Membership service provider:** One aspect of the permissioned network is that the peers are known; in Hyperledger Fabric, this is achieved across a membership service provider (MSP). The MSP ensures that the identities of the peers are known, and has the responsibility of issuing authentication and authorization credentials to the peers.

6. **Certificate authorities:** Certificate authority (CA) is an implementation of the Membership service provider, although an MSP is a default interface to cover identity management, Hyperledger Fabric enables the use of external and commercial CA.

7. **Ledger:** The ledger is a key part of a blockchain, as it stores information about any transaction that has taken place in a network. In this content, the goal is for every peer in the network to agree on the ledger states, where the instance of the ledger is deemed correct and that all parties use it as a reference.

8. **Block:** Figure 2.3 is a detailed view of the blocks in the Hyperledger Fabric framework; this will not describe over each parameter, only the ones that are deemed to further clarify the technology.

    - **Block header:** Each block header has three properties that are generated when a block is created: current block hash, previous block hash and block number.
    - **Block data:** The transactions are all simulated inside block data, where depending on specifications, blocks can have different amounts of transactions. Usually a specific size has been set for the amount of transactions, and transaction sizes. In Fabric architecture, there are three types of transactions: deploy transaction, invoke transaction, and query transaction.
    - **Block metadata:** contains information about the time in which a block is written, the block writer's certificate, public key, and signature.

9. **Transaction:** These invoke transaction in the Fabric network that change the world state. For example, a client may want to create a transaction that sends a document to another client, or a device may want to update its state in the ledger. As described in the block subsection, each block has a series of transactions, each of which has its own structure.

Consensus mechanisms Consensus mechanisms are the policies, functionalities, and algorithms that allow a distributed network to reach agreement on the fact that a block is valid, and that the ledger's current status is accurate. This means that the current ledger can be the agreed upon ledger, where everything in that ledger undergoes a policy based validation process.

Figure 2.3 has been removed from this version of the Thesis. The figure can be viewed at
https://www.researchgate.net/figure/Centralized-Multi-party-Learning-Master-Worker-Pattern-vs-Decentralized-Multi-party_fig1_330625714

FIGURE 2.3: A detailed view of a block in Hyperledger Fabric

In essence, as described in Hyperledger Fabric documentation [20], consensus mechanisms vary according to the mode of operation used, where permissionless blockchain relates to the problem of nodes being untrustworthy; as there are no policies or controls regulating participation.

The Hyperledger white paper, in addition to other articles, has introduced two properties that a consensus mechanism must satisfy [21]:

- **Safety:** Each node is guaranteed the same sequence of inputs and results in the same output on each node. Safety is an essential factor in keeping the world state of the ledger, as node agreements on the ledger results in the world state.
- **Liveness:** Each non-faulty node will eventually receive every submitted transaction, under the assumption that communication channels do not fail. Liveness establishes that all the transactions at some point reach their destination, ensuring that each transaction can undergo processing at some point.

2.1.2 Literature Review

This section presents related literature that combines usage of blockchain technology with machine learning solutions. Firstly, it reviews decentralized privacy-preserving predictive modelling on private blockchain. Next, it reviews studies that leverage blockchain technology in the healthcare industry to secure medical data privacy. Finally, it reviews research using CNNs architectures and transfer learnign technique to classify skin lesion images.

Recent advancements in machine learning methods for computer vision have resulted in the creation of state-of-the-art tools for detecting skin cancer.

Artificial Neural Networks (ANNs) [22], Support Vector Machines (SVMs) [23], Convolutional Neural Networks (CNNs) [24], hybrid deep neural networks [25], transfer learning and deep learning [26], deep CNNs [27], and segmentation [28] are some of the most widely used methods in the literature for classifying skin cancer images.

Several researchers have proposed a wide range of approaches and solutions for detecting skin cancer using dermoscopic images in recent years. These techniques range from traditional machine learning to deep learning models. Deep CNNs have been used to enhance performance in a range of applications, including visual tasks and action recognition. GoogleNet, ResNet, VGGNet, AlexNet, and EfficientNet are varied DCNN architectures that can be used in a wide range of applications. In [29], Ratul *et al.* have proposed an advanced computer-aided detection method for early detection of skin cancer. They used VGG16, VGG19, MobileNet and Inception V2 with dilated convolution on four different architectures. They trained, validated and tested the system using the HAM10000 dataset, where the architectures obtained classification accuracy of 87.42%, 85.02%, 88.22% and 89.91%.

Nasr-Esfahan *et al.* [30] have used clinical images to test deep learning. To improve the accuracy of their system, they employed a correlation for lighting and then segmented the skin moles in a preprocessing phase. A CNN was then used to extract and classify features from the enhanced and segmented images. This method produced an accuracy of 81%. Kostopoulos *et al.* [31] have suggested a computer-based analysis for palin photography. To determine the form of skin lesion, features were extracted using a Probabilistic Neural Network (PNN). Their system was able to classify 76.2% of the data.

In [32], Gessert *et al.* introduced their skin image classification system using Efficientnets, SENet, and ResNetEx WSL. Brinker *et al.* [33] have categorized skin lesion images using an improved deep learning method of a CNN architecture trained on 12,378 images. They used their model to identify 100 images, comparing the output to that of the architecture from 157 dermatologists at 12 German university hospitals. In their study, dermatologists were shown to outperform the system. In [34], Alquadah *et al.* used transfer learning, GoogleNet architecture, and gradient descent adaptive momentum learning rate (ADAM) to identify skin lesion images. Within the ISIC database, their system was evaluated to identify skin images into three categories, including seborrheic keratosis, melanoma, and benign images. This was performed on segmented and non-segmented images, with the non-segmnented classification database achieving 92.2% accuracy and non-segmented dataset achieving 89.9% accuracy.

The architectures of Inception-ResNet-V2, VGGNet, and SVM have been used to identify seven different forms of skin diseases based on images of skin. It was discovered that the Inception-ResNet-V2 architecture was more successful [35]. El-Khatib *et al.* [36] have proposed a framework using deep learning-based algorithms to diagnose skin lesions. They used pre-trained architectures such as NasNet-large, ResNet-10, and GoogleNet on large ImageNet and places 365 databases.

The research by Codella *et al.* in [37] has demonstrated the efficacy of transfer learning for skin lesion classification. Two Caffe CNN layers were used as function extractors, with their outputs being used as SVM inputs after they were pretrained on the ImageNet large scale visual recognition challenge (ILSVRC). Despite the fact that the two domain used in this transfer learning setup were extremely dissimilar, the generated finding were comparable to prior output based on hand-crafted features, demonstrating the

feasibility of extracting essential features for skin cancer detection using feature detectors trained ponphotographs of real-world images. Sparse coding was also used as an unsupervised technique in both RGB color space and grayscale, with the latter resulting in improved performance. Finally, when combinations of deep and sparse coding features were utilized (with global average across all individual performances), ensemble techniques outperformed stand-alone solutions.

Mahbod *et al.* [25] used images from the 2017 ISIC challenge as their dataset, including 2000 dermoscopic skin images with labels. They resized the images to 227x227 and 224x224 so they could be fed to the networks for compatibility. AlexNet and VGG-16 were used to extract features, and nulticlass non-linear Support Vector Machines (SVM) were used to identify them. They divided the dataset with 80% for training and 20% for testing, achieving an accuracy of 83.3% for melanoma using AlexNet and 82.7% using VGG-16.

In 2016, Shoieb *et al.* [38] used color and texture features of images to detect skin lesions and K-Means to distinguish related categories (clustering). They extracted features using a CNN and fed them into an SVM to identify skin lesions. They used three distinct datasets: the Dermatology Information System (DermIS), which contains 337 images with a binary ground truth (melanoma or non-melanome), and achieved 93.75% accuracy; DermQuest, which contains 134 images with a binary ground truth, and achieved 94.12% accuracy; and the DermNet skin Disease Atlas website, which contains 134 images with a binary ground truth, and achieved 94.12% accuracy.

Blockchain approaches have gained much interested in recent times, but mostly for their usage in cryptocurrencies like Bitcoin. However, the capabilities of this technology are not restricted to the innovations in finance. Blockchain offers a secure network architecture that may allow users to make

transactions without using a legal system, a central authority, or third parties, and thus reducing the cost of any transaction.

In addition, blockchain is a novel type of distributed ledger, since it allows for the storage of any data in the transaction metadata. Initially, Bitcoins just supported 80 bytes of metadata. Then, after several improvements, the size of its metadata was significantly increased. For instance, BigchainDB [39] does not impose a hard restriction on the number of transactions, where MultiChain [40] supports adjustable maximum metadata size per transaction. The distributed database powered by blockchain technology is often referred to as Blockchain 2.0. Ethereum is one of the most successful Blockchain 2.0 system, powered by smart contracts and DApps. The proposed method of this thesis takes advantage of transaction metadata spreading the trained learning model within a network, between distributed nodes.

Recently, the idea of blockchain 3.0 has been introduced to support applications not only in currency but also in the economy, social problems and markets. Healthcare industry, in particular, has been greatly benefited from adopting the technology for data protection and security. In [41], author has proposed a framework that integrated with decentralized blockchain to store and retrieve genomic data for the use of biomarker mining. On the other hand, Jenkins [42] takes advantage of multi-resolution blockchain to enhance the security of multi-factor authentication for the bio-mining framework. Jiasi in [43] has proposed a deep learning framework based on Blockchain to compel all participants in the network to follow the rules. This study also gurantees data privacy for each sites in the network and the auditability was provided for a whole training process. The incentive mechanism based on blockchain has been proposed in this study

Crosby *et. al.* [44] have differentiated between financial and non-financial applications that blockchain could potentially address. Accordingly, this disruptive invention has the potential to revolutionise not only the nature of relationships in finance, but also in a variety of other sectors of people's daily lives.

Al Omar *et al.* [45] have developed a blockchain-based system that addresses security and privacy concerns. They used cryptographic functions in conjunction with blockchain technology to tackle data preservation problems in their article. Patients in their system have control over whether or not thei data is shared with other users. Accordingly, this system encrypts personal health information. However, because blockchain stores user data, transaction throughput in this contest becomes poor, where the public ledger is unable to store a huge amount of data.

Blockchain technology arranges information to encourage all interested parties to verify and record transactions via a consensus algorithm. By monitoring access to confidential data and guaranteeing authorized access, blockchain technology can make sure the security of sensitive information [46]. The block-chain's PoW protocal ensures an immutable audit trail. It means that changing data becomes very difficult; as a current blockchain is linked with a previous block's hash and timestamp; thus, attackers must repeat the PoW of the target block and all subsequent blocks, and then surpass trusted sites [39].

Cirstea *et. al.* in [47] have proposed the potential for blockchain technology to revolutionize both medicine and other fields for system decentralization. Their paper showed the considerable potential of blockchain technology and how it has the potential to significantly alter all elements of receiving, transmitting, and security information. In [48], Patel *et al.* have proposed

a system for image sharing that utilizes a blockchain technique as a decentralized database to create a ledger of radiological investigations and patient-defined access to secure medical information, meeting numerous criteria for an interoperable medical system, which can be easily generalized to domains other than medical imaging. However, their proposed technique has significant drawbacks. Firstly, privacy concerns are primary considerations when utilizing blockchain technique to communicate medical data

In [49], Tsung-Ting *et al.* have proposed ModelChain as a framework for leveraging blockchain techniques for privacy-preserving learning. They have integrated online learning with a permissioned blockchain network, where they also developed a novel proof-of-information method for determining the order in which online learning processes occur. They applied privacy-preserving online machine learning algorithm in the first step. Then they used transaction's metadata to distribute the incomplete models and integrate private blockchain with online learning. Finally, they created a new proof-of-information method to be used in conjunction with the original proof-of-work consensus procedure. However, there are some limitations to this solution, such as confidentiality not being fully preserved, and transaction time might be lengthy due to the proof-of-work algorithm.

Ekblaw *et al.* [50] have suggested a new blockchain-based approach for electronic health records (EHRs). Their technology offers patients with an immutable log of their medical data, with simple access between physicians and treatment venues. The MedRec paradigm restores client agency over medical data across treatment venues and providers, arming people with the knowledge necessary to make informed healthcare decisions. The MedRec authentication log maintains health records access while giving comprehensive record inspection, data sharing, and care auditability for patients. This

model employs Ethereum smart contracts to orchestrate an access mechanism for material distributed across many storage and provider locations.

In [51], Zhang *et al.* have proposed a system for blockchain-based personal health information (PHI) sharing that maintains privay and security for the purpose of enhancing diagnosis in the e-Health system. In this framework two blockchains are presented. The permissioned blockchain holds the original PHI, while the consortium blockchain maintains records of the PHI's secure indexes. The procedure is properly constructed that the authorized physician can look for patient indexes. Additionally, the approved doctor is only permitted to examine the patient's past information and is not permitted to look for future information.

Q. Xia *et al.* [52] have presented a blockchain-based architecture for exchanging health data that adequately tackles access control problems inherent in cloud-based storage of sensitive data. Their system is built on a permissioned blockchain, only inviting and thereby verifing users. Additionally, to offer medical information with auditing, provenance, and secure data trailing for medical data, the authors use an access control mechanism and smart contracts (SCs) in another study [53] to effectively follow the behaviour of the data and revoke access to violating entities upon discovery of authorisation on data violation. In [54], to hold a metadata regarding permissions, record ownership, and data quality, a smart contract is created. The state-transition functions of the contract carry out policies, requiring only valid transactions to modify data.

Roehrs *et al.* [55] have proposed OmniPHR decentralized architecture. Their approach aims to remove persistent barriers to healthcare providers and patients adoptions of PHR. The above technique preserves the datablocks distribution feature by dispersing copies of these components across the network. This concept suggests a scalable, elastic, and interoperable architecture

enabling users to receive a single view of patient health records. With PHR data dispersed across several health institutions that patients interact with, OmniPHR offers to minimise numerous issues and hurdles to PHR adoption by offering a single picture of PHR. This concept seeks to assist individuals in taking advantage of having a single health record, as well as healthcare professionals in keeping their patients information up-to-date.

In contrast to the above studies, which focus on health data exchange, [56] and [57] address distinct concerns. [56] has proposed a blockchain-based for precision medicine and clinical trial. This above paper examines blockchain platform's design from the perspective of the medical sector, namely the precision medicine and clinical trial. Despite the fact that few research studies on key management systems for blockchain have been conducted, [57] have attempted to create key negotiation in this field. Their study utilized body sensor networks to provide a lightweight backup and recovery mechanism for health blockchain keys. This is ground-breaking work in the field of blockchain key management, as its performance is highly dependent on the hardware condition setting.

Machine learning has the potential to accelerate research and promote quality improvement activities. However, health data exchange can endanger the security of sensitive information. To protect the privacy of medical information, several strategies have been proposed to tackle predictive modeling, and effective of all is the art of transferring partially trained machine learning models. By this method, the healthcare institution can preserve its medical data while remaining up-to-date with the accurate machine learning model. For instance, GLORE [58] uses horizontally partitioned data to build logistic regression models while VERTIGO [59] deals with vertically partitioned data.

Secure multi-party computing (SMC) is a common methodology, where each participant employs a series of cryptographic algorithms and the oblivious transfer method to jointly compute a function using their private data [60]. This type of mechanism can ensure that none of the parties can learn anything about the data of the other. Cryptographic approaches, in contrast, are computationally costly for data owners due to the repetitive encyption and decryption processes.

Xuhui *et al.* [61] have presented a learning model for the permissioned blockchain. Additionally, they have developed a privacy-preserving gradient computing based on differential privacy. Their framework has three stages: blockchain initiation; local gradient computation; and global gradient computation. Using the Ethereum framework, their model is trained and evaluated on different datasets: synthetic dataset; Wisconsin breast cancer dataset; and the MNIST dataset. However, because their method is based on proof-of-work, each iteration takes a long time to complete.

Payemnt *et. al.* [62] have proposed a privacy preserving logistic regression protocol. They proposed many protocols for privacy preserving machine learning using stochastic gradient descent (SGD). Their protocol implements on both online and offline phases. However, the data privacy is not addressed sufficiently. In [63], Yan et. al. have proposed distributed autonomous online learning method that computes local subgradients and distributes parameter vectors across network nodes. However, this machine learning algorithm was developed using a centralized network architecture, which introduces security vulnerabilities such as a single-point-of-failure.

Kim et. al. [64] have proposed a privacy-preserving distributed machine learning (DML). They developed an error-based technique and a private stochastic gradient descent algorithm. Their framework is divided into

three phases: simulation phase, ordering phase and execution phase. Their model is based on permission blockchain, so identity privacy is secured. Nevertheless, the privacy infringement caused data leaking through the training process has still not been considered thoroughly.

There are many research studies on using blockchain and machine learning together, however they have not focused sufficiently on protecting data privacy. The authors in [49] and [61] have presented a permissionless blockchain machine learning approach. In any environment and/or situation, a public blockchain seeks to be a completely decentralized and scalable network. However, they these studies have not adequately addressed security, privacy, and efficiency issues, where their proposed algorithms contain structural restrictions and vulnerabilities. Exchange-based learning has been proposed by [49]. This involves the next global weight being determined by a participant's local weight with lowest error in each iteration. This trade is based on the suggested consensus algorithm, which is based on PoW. Even if they employ the public blockchain functionality, it merely presumes that the network is impenetrable to the outside world. They also fail to address data privacy concerns, where their consensus is inefficient, since PoW-based agreement requires a significant amount of time for each iteration of learning. Paper [61] uses aggregation-based learning to calculate a global weight, which is calculated by combining numerous local gradients in each iteration. Additionally, they devise a gradient computing that preserves privacy is based on differential privacy (DP). Because their technique is dependent on PoW, each iteration takes a long time.

Besides these distributed privacy-preserving algorithms, there are many "online" learning approaches that can be changed in a consecutive order (in

contrast to the other "batch" algorithms). For example: the Distributed Autonomous Online Learning [65] and EXPLORER [66], whose online machine learning algorithms aim to update models on the blockchain sequentially.

Nevertheless, all the algorithms mentioned above rely on the centralized architecture by design. Therefore, any model update based on batches or online fashion still makes these algorithms exposed to vulnerable issues associated with a single-point-of-failure. In contrast, our proposed method architecture is based on blockchain, which is decentralization. Thus, it can improve further robustness and security of the system.

It can be noted that other research studies also covered the decentralized architecture for enhancing the data-parallelism in machine learning, one not able solution is the Map-Reduce technology. However, instead of aiming at analysing and providing security layers for data protection, these procedures have mainly focused on parallelization methods to increase computational power, thus, reducing the time taken for training and inference.

2.1.3 Research questions

As discussed above, health information exchange (HIE) can assist organizations in improving healthcare quality and patient outcomes. However, data privacy concern arise as an urgent task that needs to be solved. Numerous studies have examined different approaches and algorithms to solve this problem. Nevertheless, to date, privacy and security concerns have not been adequately addressed. Accordingly, this leads to the research questions described below:

1. *Is it possible to develop an online learning platform for machine learning?*

In practice, machine learning is frequently utilized to generate predictive models for applications. When trained on a vast quantity of data collected from a variety of sources, these models become increasingly accurate. This enormous data gathering; however, creates privacy issues. This thesis presents an online learning platform for machine learning that enables many parties to collaboratively develop an accurate model for a given target without revealing their input dataset.

2. *Is it possible to develop a security mechanism using a private blockchain network to secure the healthcare data?*

   This thesis leverages permissioned blockchain to ensure the integrity and privacy of all medical data during its training process. Our framework uses the information-bidding mechanism that allows hospitals in the network to exchange their models without exposing sensitive information and continuously upgrading their model.

Chapter 3 is not available in this version of the Thesis as it is intended for publication.

Chapter 4 is not available in this version of the Thesis as it is intended for publication.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this thesis, we introduced a private blockchain technique to increase the security and robustness of distributed online learning. The thesis presented a framework for combining machine learning models and a blockchain network. By training the model using a transfer learning technique and broadcasting the model parameters using a Hyperledger Fabric blockchain network, our privacy-preserving online learning framework achieves promising results in the form of improved prediction accuracy, especially when the training datasets are small at local sites.

The thesis showed that the performance of the model improved when it went through all datasets in the network. The proposed framework can generate more predictive models to help clinical/ biomedical/ genomics studies with future assessments and modifications.

5.2 Skin Lesion classification using Transfer Learning

Despite the availability of imaging and detection procedures for melanoma, such as dermoscopy, automatic recognition remains problematic due to the difficulties of segmenting precise lesion areas, similarities between melanoma and non-melanoma lesions, and the wide range of skin disorders. Aside from these issues, finding medical images while maintaining patient's privacy is difficult. We employed our proposed model for image classification using transfer learning algorithm with pre-trained ImageNet dataset weights. Our proposed solution allows multiple sites to collaborate and share model parameters based on local datasets. This ultimately leads to better detection accuracies while preserving the security and privacy of health records.

5.3 Future research direction

In our current model, information bidding works as a "hill climbing" mechanism. One can consider adapting it to other natural process, for example, simulated annealing. This can potentially reduce the time required for all hospitals to send back their models. One can consider decentralization of the the decision making server to get our framework fully decentralized.

In order to create decentralized learning systems that can train on private dataset, incentive systems for contributing data will be crucial. In non-decentralized learning systems, participants were encouraged to share data with centralized services in exchange for storage and better machine learning models based on their data. Now with decentralized and distributed learning, the medical data of all hospitals are stored in their dataset and the model

gets updated based on their local dataset without expose their data. More studies and simulations should be done to identify incentive systems that reward contributors while boosting a model performance and preventing bias. This is especially significant in healthcare, where a model's ability to learn from experiences of other patients might save a patient's life. In this work, we used blockchain for maintaining privacy of health records. Other privacy and security measures can be investigated in future works for developing more robust systems.

Bibliography

[1] Aaron Boddy et al. “An Investigation into Healthcare-Data Patterns”. In: *Future Internet* 11.2 (2019), p. 30.

[2] healthIT.gov. “Health IT and Health Information Exchange Basics”. In: *healthIT.gov* (2018).

[3] Francisca Adoma Acheampong. “Big Data Machine Learning and the BlockChain Technology: An Overview”. In: *International Journal of Computer Applications* 975 (2018), p. 8887.

[4] Xuhui Chen et al. “Cost-sensitive deep active learning for epileptic seizure detection”. In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2018, pp. 226–235.

[5] George D Magoulas and Andriana Prentza. “Machine learning in medical applications”. In: *Advanced course on artificial intelligence*. Springer. 1999, pp. 300–307.

[6] Vassilios S Verykios et al. “State-of-the-art in privacy preserving data mining”. In: *ACM Sigmod Record* 33.1 (2004), pp. 50–57.

[7] Jakub Konečnỳ et al. “Federated learning: Strategies for improving communication efficiency”. In: *arXiv preprint arXiv:1610.05492* (2016).

[8] Dylan Yaga et al. “Blockchain technology overview”. In: *arXiv preprint arXiv:1906.11078* (2019).

[9] Massimo Di Pierro. "What is the blockchain?" In: *Computing in Science & Engineering* 19.5 (2017), pp. 92–95.

[10] ForgeRock. "U.S. Consumer Data Breach Report 2019". In: *ForgeRock* (2019).

[11] Nir Kshetri. "Blockchain and Electronic Healthcare Records [Cybertrust]". In: *Computer* 51.12 (2018), pp. 59–63.

[12] J. K. Cohen. "Healthcare ranks 8th out of 18 industries for data security performance, report says". In: *Becker's Hospital Review* (2019).

[13] SAMER Hijazi, RISHI Kumar, and CHRIS Rowen. "Using Convolutional Neural Networks for Image Recognition. 2015". In: *Dostupné z:< https://ip. cadence. com/uploads/901/cnn_wp-pdf* (2020).

[14] Martin Thoma. "Analysis and optimization of convolutional neural network architectures". In: *arXiv preprint arXiv:1707.09725* (2017).

[15] Alexandru Stanciu. "Blockchain based distributed control system for edge computing". In: *2017 21st International Conference on Control Systems and Computer Science (CSCS)*. IEEE. 2017, pp. 667–671.

[16] Alysson Bessani, João Sousa, and Marko Vukolić. "A byzantine fault-tolerant ordering service for the hyperledger fabric blockchain platform". In: *Proceedings of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers*. 2017, pp. 1–2.

[17] Andreas M Antonopoulos. *Mastering Bitcoin: unlocking digital cryptocurrencies*. " O'Reilly Media, Inc.", 2014.

[18] Oscar Novo. "Blockchain meets IoT: An architecture for scalable access management in IoT". In: *IEEE Internet of Things Journal* 5.2 (2018), pp. 1184–1195.

[19] Vikram Dhillon, David Metcalf, and Max Hooper. "The hyperledger project". In: *Blockchain enabled applications*. Springer, 2017, pp. 139–149.

[20] *Hyperledger Fabric documentation*. 2020. URL: https://hyperledger-fabric.readthedocs.io/en/release-2.2/ledger/ledger.html.

[21] Ingo Weber et al. "Untrusted business process monitoring and execution using blockchain". In: *International Conference on Business Process Management*. Springer. 2016, pp. 329–347.

[22] Uğur Fidan, İsmail Sarı, and Raziye Kübra Kumrular. "Classification of skin lesions using ANN". In: *2016 Medical Technologies National Congress (TIPTEKNO)*. IEEE. 2016, pp. 1–4.

[23] HR Mhaske and DA Phalke. "Melanoma skin cancer detection and classification based on supervised and unsupervised learning". In: *2013 International conference on Circuits, Controls and Communications (CCUBE)*. IEEE. 2013, pp. 1–5.

[24] Titus Josef Brinker et al. "Skin cancer classification using convolutional neural networks: systematic review". In: *Journal of medical Internet research* 20.10 (2018), e11936.

[25] Amirreza Mahbod et al. "Skin lesion classification using hybrid deep neural networks". In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 1229–1233.

[26] Khalid M Hosny, Mohamed A Kassem, and Mohamed M Foaud. "Skin cancer classification using deep learning and transfer learning". In: *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*. IEEE. 2018, pp. 90–93.

[27] Marwan Ali Albahar. "Skin lesion classification using convolutional neural network with novel regularizer". In: *IEEE Access* 7 (2019), pp. 38306–38313.

[28] Mobeen ur Rehman et al. "Classification of skin lesion by interference of segmentation and convolotion neural network". In: *2018 2nd International Conference on Engineering Innovation (ICEI)*. IEEE. 2018, pp. 81–85.

[29] Md Aminur Rab Ratul et al. "Skin lesions classification using deep learning based on dilated convolution". In: *BioRxiv* (2020), p. 860700.

[30] Ebrahim Nasr-Esfahani et al. "Melanoma detection by analysis of clinical images using convolutional neural network". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, pp. 1373–1376.

[31] Spiros A Kostopoulos et al. "Adaptable pattern recognition system for discriminating Melanocytic Nevi from Malignant Melanomas using plain photography images from different image databases". In: *International journal of medical informatics* 105 (2017), pp. 1–10.

[32] Nils Gessert et al. "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data". In: *MethodsX* 7 (2020), p. 100864.

[33] Titus J Brinker et al. "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task". In: *European Journal of Cancer* 113 (2019), pp. 47–54.

[34] Ali Mohammad Alqudah, Hiam Alquraan, and Isam Abu Qasmieh. "Segmented and non-segmented skin lesions classification using transfer learning and adaptive moment learning rate technique using pretrained convolutional neural network". In: *Journal of Biomimetics, Biomaterials and Biomedical Engineering*. Vol. 42. Trans Tech Publ. 2019, pp. 67–78.

[35] Shetu Rani Guha and SM Rafizul Haque. "Performance comparison of machine learning-based classification of skin diseases from skin lesion

images". In: *International conference on communication, computing and electronics systems*. Springer. 2020, pp. 15–25.

[36] Hassan El-Khatib, Dan Popescu, and Loretta Ichim. "Deep learning–based methods for automatic diagnosis of skin lesions". In: *Sensors* 20.6 (2020), p. 1753.

[37] Noel Codella et al. "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images". In: *International workshop on machine learning in medical imaging*. Springer. 2015, pp. 118–126.

[38] Doaa A Shoieb, Sherin M Youssef, and Walid M Aly. "Computer-aided model for skin diagnosis using deep learning". In: *Journal of Image and Graphics* 4.2 (2016), pp. 122–129.

[39] Trent McConaghy et al. "Bigchaindb: a scalable blockchain database". In: *white paper, BigChainDB* (2016).

[40] Gideon Greenspan. "Multichain private blockchain-white paper". In: *URl: http://www. multichain. com/download/MultiChain-White-Paper. pdf* (2015).

[41] Kevin Judd McKernan. "The chloroplast genome hidden in plain sight, open access publishing and anti-fragile distributed data sources". In: *Mitochondrial DNA Part A* 27.6 (2016), pp. 4518–4519.

[42] Jeffrey Jenkins et al. "Bio-mining for biomarkers with a multi-resolution block chain". In: *Independent Component Analyses, Compressive Sampling, Large Data Analyses (LDA), Neural Networks, Biosystems, and Nanoengineering XIII*. Vol. 9496. International Society for Optics and Photonics. 2015, 94960N.

[43] Jiasi Weng et al. "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive". In: *IEEE Transactions on Dependable and Secure Computing* (2019).

[44] Michael Crosby et al. "Blockchain technology: Beyond bitcoin". In: *Applied Innovation* 2.6-10 (2016), p. 71.

[45] Abdullah Al Omar et al. "Medibchain: A blockchain based privacy preserving platform for healthcare data". In: *International conference on security, privacy and anonymity in computation, communication and storage*. Springer. 2017, pp. 534–543.

[46] Guy Zyskind, Oz Nathan, et al. "Decentralizing privacy: Using blockchain to protect personal data". In: *2015 IEEE Security and Privacy Workshops*. IEEE. 2015, pp. 180–184.

[47] Andrei Cirstea et al. "Blockchain technology applied in health the study of blockchain application in the health system (II)". In: *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. IEEE. 2018, pp. 1–4.

[48] Vishal Patel. "A framework for secure and decentralized sharing of medical imaging data via blockchain consensus". In: *Health informatics journal* 25.4 (2019), pp. 1398–1411.

[49] Tsung-Ting Kuo and Lucila Ohno-Machado. "Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks". In: *arXiv preprint arXiv:1802.01746* (2018).

[50] Ariel Ekblaw et al. "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.

[51] Aiqing Zhang and Xiaodong Lin. "Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain". In: *Journal of medical systems* 42.8 (2018), pp. 1–18.

[52] Qi Xia et al. *BBDS: blockchain-based data sharing for electronic medical records in cloud environments, Information, 8 (2)(2017)*. 2017.

[53] QI Xia et al. "MeDShare: Trust-less medical data sharing among cloud service providers via blockchain". In: *IEEE Access* 5 (2017), pp. 14757–14767.

[54] Asaph Azaria et al. "Medrec: Using blockchain for medical data access and permission management". In: *2016 2nd International Conference on Open and Big Data (OBD)*. IEEE. 2016, pp. 25–30.

[55] Alex Roehrs, Cristiano André Da Costa, and Rodrigo da Rosa Righi. "OmniPHR: A distributed architecture model to integrate personal health records". In: *Journal of biomedical informatics* 71 (2017), pp. 70–81.

[56] Zonyin Shae and Jeffrey JP Tsai. "On the design of a blockchain platform for clinical trial and precision medicine". In: *2017 IEEE 37th international conference on distributed computing systems (ICDCS)*. IEEE. 2017, pp. 1972–1980.

[57] Huawei Zhao et al. "Lightweight backup and efficient recovery scheme for health blockchain keys". In: *2017 IEEE 13th international symposium on autonomous decentralized system (ISADS)*. IEEE. 2017, pp. 229–234.

[58] Yuan Wu et al. "G rid Binary LO gistic RE gression (GLORE): building shared models without sharing data". In: *Journal of the American Medical Informatics Association* 19.5 (2012), pp. 758–764.

[59] Yong Li et al. "Vertical grid logistic regression (vertigo)". In: *Journal of the American Medical Informatics Association* 23.3 (2016), pp. 570–579.

[60] Jaideep Vaidya et al. "Privacy-preserving decision trees over vertically partitioned data". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 2.3 (2008), pp. 1–27.

[61] Xuhui Chen et al. "When machine learning meets blockchain: A decentralized, privacy-preserving and secure design". In: *2018 IEEE International Conference on Big Data (Big Data)*. IEEE. 2018, pp. 1178–1187.

[62] Payman Mohassel and Yupeng Zhang. "Secureml: A system for scalable privacy-preserving machine learning". In: *2017 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2017, pp. 19–38.

[63] Feng Yan et al. "Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties". In: *IEEE Transactions on Knowledge and Data Engineering* 25.11 (2012), pp. 2483–2493.

[64] Hyunil Kim et al. "Efficient privacy-preserving machine learning for blockchain network". In: *IEEE Access* 7 (2019), pp. 136481–136495.

[65] Stephen E Fienberg et al. ""Secure" log-linear and logistic regression analysis of distributed databases". In: *International Conference on Privacy in Statistical Databases*. Springer. 2006, pp. 277–290.

[66] Shuang Wang et al. "Expectation propagation logistic regression (explorer): distributed privacy-preserving online model learning". In: *Journal of biomedical informatics* 46.3 (2013), pp. 480–496.

[67] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[68] Christian Szegedy et al. "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

[69] Mingxing Tan and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6105–6114.

[70] D. Parkin D. Mesher and Sasieni. "Cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010". In: *British journal of cancer 105* (2011).

[71] Akila Victor and Muhammad Rukunuddin Ghalib. "Detection of Skin Cancer Cells–A Review". In: *Research Journal of Pharmacy and Technology* 10.11 (2017), pp. 4093–4098.

[72] Worl Health Organization. "World, Health, and Organization. How common is skin cancer?" In: (access on 2021). URL: https://www.who.int/news-room/q-a-detail/radiation-ultraviolet-(uv)-radiation-and-skin-cancer.

[73] Ammara Masood and Adel Ali Al-Jumaily. "Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms". In: *International journal of biomedical imaging* 2013 (2013).

[74] Howard K Koh. "Melanoma screening: focusing the public health journey". In: *Archives of dermatology* 143.1 (2007), pp. 101–103.

[75] American Cancer Society. "Cancer Facts Figures". In: (2018). URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf.

[76] Jose Luis García Arroyo and Begoña García Zapirain. "Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis". In: *Computers in biology and medicine* 44 (2014), pp. 144–157.

[77] Melanoma Institute Australia. "Melanoma facts and statistics". In: (access in 2021). URL: https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/cancer-summary-data-visualisation.

[78] Australia Government. "Melanoma of the skin statistics". In: (2020). URL: https://www.canceraustralia.gov.au/affected-cancer/cancer-types/melanoma/melanoma-skin-statistics.

[79] Australian goverment. "Melanoma of the skin statistics". In: (2020). URL: https://www.canceraustralia.gov.au/affected-cancer/cancer-types/melanoma/melanoma-skin-statistics.

[80] Tryan Aditya Putra, Syahidah Izza Rufaida, and Jenq-Shiou Leu. "Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation". In: *IEEE Access* 8 (2020), pp. 40536–40546.

[81] Noel CF Codella et al. "Deep learning ensembles for melanoma recognition in dermoscopy images". In: *IBM Journal of Research and Development* 61.4/5 (2017), pp. 5–1.

[82] Darrell S Rigel. "Non-invasive gene expression testing to rule out melanoma". In: (2018).

[83] B Aika Shoo, Richard W Sagebiel, and Mohammed Kashani-Sabet. "Discordance in the histopathologic diagnosis of melanoma at a melanoma referral center". In: *Journal of the American Academy of Dermatology* 62.5 (2010), pp. 751–756.

[84] Laura K Ferris et al. "Real-world performance and utility of a noninvasive gene expression assay to evaluate melanoma risk in pigmented lesions". In: *Melanoma research* 28.5 (2018), pp. 478–482.

[85] R Marks. "Epidemiology of melanoma: Clinical dermatology• Review article". In: *Clinical and Experimental Dermatology: Clinical dermatology* 25.6 (2000), pp. 459–463.

[86] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. "Cancer statistics, 2016". In: *CA: a cancer journal for clinicians* 66.1 (2016), pp. 7–30.

[87] Mohammad Ali Kadampur and Sulaiman Al Riyaee. "Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images". In: *Informatics in Medicine Unlocked* 18 (2020), p. 100282.

[88] Kaggle. "The ISIC 2020 Challenge Dataset". In: (2020).

[89] Andre Esteva et al. "Dermatologist-level classification of skin cancer with deep neural networks". In: *nature* 542.7639 (2017), pp. 115–118.

[90] Ulzii-Orshikh Dorj et al. "The skin cancer classification using deep convolutional neural network". In: *Multimedia Tools and Applications* 77.8 (2018), pp. 9909–9924.

[91] Lisa Torrey and Jude Shavlik. "Transfer learning". In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

[92] Nidhi Ranjan et al. "Hierarchical approach for breast cancer histopathology images classification". In: (2018).

[93] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical image analysis* 42 (2017), pp. 60–88.

[94] David R Kaeli et al. *Heterogeneous computing with OpenCL 2.0*. Morgan Kaufmann, 2015.

[95] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

[96] Jason Brownlee. "A gentle introduction to transfer learning for deep learning". In: *Machine Learning Mastery* (2017).

[97] Shuo Feng, Huiyu Zhou, and Hongbiao Dong. "Using deep neural network with small dataset to predict material defects". In: *Materials & Design* 162 (2019), pp. 300–310.

[98] Hong-Wei Ng et al. "Deep learning for emotion recognition on small datasets using transfer learning". In: *Proceedings of the 2015 ACM on international conference on multimodal interaction*. 2015, pp. 443–449.

[99] Wei Zhao. "Research on the deep learning of the small sample data based on transfer learning". In: *AIP Conference Proceedings*. Vol. 1864. 1. AIP Publishing LLC. 2017, p. 020018.

[100] Mark Sandler et al. "Mobilenetv2: Inverted residuals and linear bottlenecks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 4510–4520.

[101] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. "A survey of transfer learning". In: *Journal of Big data* 3.1 (2016), pp. 1–40.

[102] Maxime Oquab et al. "Learning and transferring mid-level image representations using convolutional neural networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1717–1724.

[103] Agnieszka Mikołajczyk and Michał Grochowski. "Data augmentation for improving deep learning in image classification problem". In: *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE. 2018, pp. 117–122.

[104] Nitish Srivastava et al. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[105] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *International conference on machine learning*. PMLR. 2015, pp. 448–456.

[106] Mohamed A Kassem, Khalid M Hosny, and Mohamed M Fouad. "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning". In: *IEEE Access* 8 (2020), pp. 114822–114832.

[107] Amol S Navathe and Patrick H Conway. "Optimizing health information technology's role in enabling comparative effectiveness research". In: *Am J Manag Care* 16 (2010), p. 12.

[108] Paul Wicks et al. "Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm". In: *Nature biotechnology* 29.5 (2011), pp. 411–414.

[109] Joy M Grossman et al. "Creating sustainable local health information exchanges: can barriers to stakeholder participation be overcome?" In: (2008).

[110] Naomi J. Alpern and Robert J. Shimonski. "CHAPTER 1 - Network Fundamentals". In: *Eleventh Hour Network+*. Ed. by Naomi J. Alpern and Robert J. Shimonski. Boston: Syngress, 2010, pp. 1–18. ISBN: 978-1-59749-428-1. DOI: https://doi.org/10.1016/B978-1-59749-428-1.00003-5. URL: https://www.sciencedirect.com/science/article/pii/B9781597494281000035.

[111] Conner Fromknecht, Dragos Velicanu, and Sophia Yakoubov. "A Decentralized Public Key Infrastructure with Identity Retention." In: *IACR Cryptol. ePrint Arch.* 2014 (2014), p. 803.

[112] Loi Luu et al. "Scp: A computationally-scalable byzantine consensus protocol for blockchains". In: *www.weusecoins. com/assets/pdf/library/SCP* 20.20 (2015), p. 2016.

[113] Imran Bashir. *Mastering blockchain*. Packt Publishing Ltd, 2017.

[114] Christian Cachin et al. "Architecture of the hyperledger blockchain fabric". In: *Workshop on distributed cryptocurrencies and consensus ledgers*. Vol. 310. 4. Chicago, IL. 2016.

[115] Per-Erik Danielsson. "Euclidean distance mapping". In: *Computer Graphics and image processing* 14.3 (1980), pp. 227–248.

[116] Satoshi Nakamoto. "Bitcoin: A peer-to-peer electronic cash system Bitcoin: A Peer-to-Peer Electronic Cash System". In: *Bitcoin. org. Disponible en https://bitcoin. org/en/bitcoin-paper* (2009).

[117] Gavin Wood et al. "Ethereum: A secure decentralised generalised transaction ledger". In: *Ethereum project yellow paper* 151.2014 (2014), pp. 1–32.

[118] Elli Androulaki et al. "Hyperledger fabric: a distributed operating system for permissioned blockchains". In: *Proceedings of the thirteenth EuroSys conference*. 2018, pp. 1–15.

[119] Vitalik Buterin et al. "A next-generation smart contract and decentralized application platform". In: *white paper* 3.37 (2014).