

Identifying Dark Web Hidden Services with Novel Image Classes Using CNN and Quantum Transfer Learning

Ashwini Dalvi*

Department of Computer Engineering, Veermata Jijabai Technological Institute, India, Mumbai, 400019, India

E-mail: aadalvip19@ce.vjti.ac.in

ORCID iD: <https://orcid.org/0000-0001-9015-457X>

*Corresponding Author

Soham Bhoir

Department of Information Technology, K J Somaiya College of Engineering, Vidyavihar, 400077, India

E-mail: soham.bhoir@somaiya.edu

ORCID iD: <https://orcid.org/0000-0002-4776-8602>

Akansha Singh

Department of Computer Engineering, Veermata Jijabai Technological Institute, India, Mumbai, 400019, India

E-mail: m20@ce.vjti.ac.in

Irfan Siddavatam

Department of Information Technology, K J Somaiya College of Engineering, Vidyavihar, 400077, India

E-mail: irfansiddavatam@somaiya.edu

ORCID iD: <https://orcid.org/0000-0002-0858-3154>

Sunil Bhirud

Department of Computer Engineering, Veermata Jijabai Technological Institute, India, Mumbai, 400019, India

E-mail: sgbhirud@ce.vjti.ac.in

Received: 23 October, 2023; Revised: 08 December, 2023; Accepted: 09 December, 2024; Published: 08 April, 2025

Abstract: The dark web is an overwhelming and mysterious place that comprises hidden services. Dark web hidden services contain illegal or offensive content. Hidden services are not accessible through regular search engines or browsers and can only be accessed via specific software. The proposed work aims to identify these hidden services by analyzing their associated images and text data. Doing so, one can better understand the types of activities on the dark web and what kind of content is available. First, a dark web crawler is developed to collect dark web services. Images are then manually classified into four categories: Cards, Devices, Hackers, and Money. Next, preprocessing the collected dataset removed irrelevant images, and a Convolutional Neural Network (CNN) was trained to identify new dark web image classes. Finally, quantum Transfer Learning (QTL) improved the model's performance. The proposed work goes beyond conventional methods of categorizing datasets by including new categories of image classes of dark web hidden services that have not been considered before. Also, the work examines image data and related text to establish a strong correlation between them. The proposed approach will provide insights into the dark web hidden service by confirming the relationship between the image and text data of the respective hidden-services.

Index Terms: Dark Web, Image classification, CNN Model, Quantum Transfer Learning, TF-IDF

1. Introduction

An onion service, a hidden service, is a website hosted on the Tor dark web. Hidden services have the advantage of hiding their IP addresses from the outside world, making tracking them difficult. In addition, the Tor network provides hidden services instead of a traditional domain name. Therefore, hidden services are only accessible through the

Tor network using .onion domain names. A .onion domain uses a pseudo-top-level domain, not a top-level domain like .com or .org. Hidden services are typically accessible by knowing their onion domain names (URLs), which comprise 16 random characters. Upon entering the .onion domain name, a request is encrypted and routed through the Tor network relays until it reaches the server.

Since traffic to and from the hidden site is encrypted and routed through the Tor network, locating the server hosting the site or identifying its owner is challenging [26]. Encrypted and anonymous communication on the dark web makes it an ideal platform for people who want to remain anonymous while browsing the web. However, the dark web is infamous as an online platform notorious for committing illicit activities such as selling drugs, firearms, and child sexual abuse material (CSAM) [22, 7]. It also serves as a platform for cybercriminals to conduct their activities [20]. For extremist groups, the dark web provides a platform for communication, coordination, and recruitment [33]. Several of these activities may significantly threaten public safety and national security.

With the increasing growth of dark web data, it is becoming increasingly important to classify it to make it more useful and understandable for humans and machines. However, searching thousands of dark web services for their sensitive nature would take extensive time and resources. Artificial Intelligence (AI) and Machine learning (ML) algorithms can be used to analyze large amounts of data on the dark web to identify patterns and potentially detect criminal activity [27]. Through deep learning, algorithms can be used to uncover hidden services and activities that are taking place on the Tor network. By analyzing these data points, law enforcement agencies can gain helpful insights and take necessary action against criminals in this space.

Thus, automatically identifying and classifying dark web services based on their images and content allows researchers and law and enforcement agencies (LEAs) to identify and track the potentially illegal activity on the dark web. Unfortunately, quality data for dark web data classification with ML algorithms are lacking. Also, data on the dark web is often incomplete and unreliable because of the difficulty of accessing hidden services. For example, in most studies of the Tor dark web, websites were manually categorized, and then automated crawlers were trained using those categories. However, this method cannot accomplish law enforcement tasks that require specific categories [8]. On the other hand, studies on illegal online activity tend to focus on specific types of crimes like terrorism.

Further, several research studies analyzed dark web services with text classification approaches with customized data sets. For example, for dark web illicit activities classification, machine learning algorithms were trained using laws and regulations related to various types of illegal activity [18]. The dark web crawler Crawlbot's ANN-based classifier was trained with Pornography Text Dataset to improve accuracy and efficiency in detecting pornographic content on the dark web [30]. However, over 75% of dark market listings contain image data, highlighting the significance of considering image data for analysis [19].

Identifying illegal or illicit activities such as illicit drug selling, child pornography, or other criminal activity would be efficient by observing the visual content of a hidden service. Accordingly, some studies analyzed dark web image content. For example, the TOR Image Categories (TOIC) dataset with five illegal classes was presented and categorized using the Bag of Visual Words model, merging Edge-SIFT and dense SIFT features [14]. However, most are restricted to specific domains, such as child abuse or weapons, leaving out images associated with other categories.

Most automated techniques with natural language processing (NLP) and text analysis analyze and process textual data. However, these techniques do not enable the analysis of non-textual objects, such as images, videos, or audio recordings, as effectively as they do for textual objects. However, computer vision, which develops algorithms and techniques to analyze and interpret visual data, such as images and videos, has progressed. Integrating computer vision and natural language processing techniques will enable a more comprehensive multimedia data analysis.

In the presented study, authors proposed using image data to label dark web services. This research aims to determine hidden services by analyzing their images by correlating textual data of respective hidden services. The contributions of the proposed research are as follows:

- Identifying four new image classes of hidden services - Card, Device, Hacker, and Money As a result of this contribution, these four image classes have been identified and labeled, which can be used as training data for image classification algorithms.
- Extending the concept of transfer learning to hybrid classical-quantum neural networks for image classification. The proposed hybrid model used transfer learning techniques to learn classical and quantum neural networks to outperform better than traditional CNN.
- Confirming the correlation between textual data and image data of a particular hidden service Using textual and image data together allowed authors to confirm gathered insights into hidden service. In addition, the authors hypothesized that new insights and discoveries might emerge that would not otherwise be possible by analyzing text and image data types.

The paper is organized as follows: Section 2 reviews the existing literature on the topic, highlighting the contributions and limitations of previous studies. Section 3 presents the methodology, including image classification using the CNN model, quantum transfer learning models, and text feature engineering approaches. Section 4 provides the results and discussion. Section 5 outlines the limitations of the current approach. Finally, Section 6 concludes the paper and suggests directions for future research.

2. Review of Previous Studies

Dark Web sites are secret parts of the internet that are not indexed by search engines and are available only through specialized software or configurations. Hackers and cybercriminals often use the platform to sell stolen data, traffic drugs, and engage in other illegal activities. Dark web content is analyzed by researchers for Cyber Threat Intelligence (CTI) in order to identify these illegal activities [3]. In addition to manual analysis, automated tools can identify patterns and anomalies in the data using machine learning and natural language processing algorithms. The dark web has been scraped and mined for threat intelligence purposes using web scraping and data mining techniques.

The data repositories and crawling are preferred techniques to collect dark web hidden services for labeling and classification. For example, researchers created a dataset containing Darknet domain names referred to as “Darknet Use Text Addresses (DUTA)” [2]. The text data was classified using three different classifiers (SVM, LR, and NB) and two text representation methods (TF-IDF and BOW). The tool BlackWidow collects HTML-based forum data to extract unstructured dark web data [28]. The BlackWidow tool collected and analyzed large amounts of unstructured dark web data. A study of seven popular Deep and Dark Web services was conducted in three different languages with the help of BlackWidow.

Researchers discussed methods to label dark web services. Researchers extended the scope of content labeling of hidden services with a keyword extraction method where relevant keywords were extracted from hidden service content [9]. In addition, the graph degeneracy method was applied to identify their most significant ones. To overcome the limitation of the requirement of labeled content to identify and characterize the dark web, which is often expensive and difficult to obtain, researchers proposed an unsupervised model for identifying and describing dark web forums combining clustering and decision tree algorithms [25].

Table 1. Dark Web Hidden Service Identification with Images

Sr No.	Referred Dataset	Contribution	Limitations of the study
1	Darknet Market Archives 2013-2015 (dnmarchives) [19, 6]	The research focused on the effectiveness of using image hashing for identifying similar images between dark marketplaces for vendor identification	The scope of research is limited to vendor profiling
2	Authors crawled the dark web to create TOIC (TOR Image Categories) [14]	The research classified images into the following classes: • Drugs, • Weapons, • Money, • Personal ID and • Credit Cards	Requirement of improvement in the efficiency of Compass Radius Estimation for Image Classification (CREIC) algorithm closer to the Ideal Radius Selection algorithm
3	TOIC (TOR Image Categories) [13]	Classification of images with SAKF - Semantic Attention Keypoint Filtering algorithm to detect foreground objects of interest	The images extracted from the dark web do not always show an object of interest in the ideal situation (without blur or similar distorting effects in images) for the proposed method.
4	TOIC (TOR Image Categories) and other public datasets* [12]	On a custom dataset, CNN features MobileNet v1, Resnet50, and BoVW utilizing dense SIFT descriptors compared against SAKF. SAKF outperformed alternative approaches.	Seeking foreground objects with the removal of background noise could be challenging in dark web crawled images
5	DUSI (Darknet Usage Service Images) [4]	Perceptual hashing (pHash) and Bag of Visual Words have been used to recognize the service domains on TOR dark net. The research classified images into the following classes: • Cryptocurrency • Cryptolockers • Hosting • Locked • Personal ID	There was no mention of text data correlation with images. The dataset used is not available publicly.
6	circl-ail-dataset-01 [24]	Analyzing data is divided into two modalities. First, using Convolutional Neural Networks (CNN) to process image-based data and incorporate Grad-CAM to explain the findings. The second modality analyzed text-based.	Work done on the existing data set.

Text-based data scraped from the dark web has been used for most related work. However, artifacts from the dark web that rely on images rather than text have not fully been considered in studies on hidden services. In addition, the dark web could be better monitored and prosecuted if law enforcement agencies also analyzed image artifacts. For example, the study suggested that analyzing product photos can help connect multiple accounts of the same darknet vendor [34]. Also, researchers employed image processing techniques to automate systems to recognize and decipher the characters in the CAPTCHA image to enter the dark web marketplace [35].

Researchers used image classification to identify violent extremist organizations that post real-time visual propaganda on social networks [16]. Video or image propaganda promoting violent extremism was to be identified using surveillance techniques to monitor online activity.

For Tor dark web image research, researchers introduced DUSI-2K, a dataset including 2500 snapshots of Tor domains [5]. The dataset is divided into illicit services and not illicit services images. Furthermore, researchers introduced a robust image hashing scheme called F-DNS to classify Tor domains. Perceptual Hashing did not require training in the model; thus, the time investment is less than Bag of Visual Word (BOVW). Also, Perceptual Hashing provided higher accuracy than BOVW for the given dataset.

Table I offers summarized findings of dark web hidden service identification with images.

The analysis of dark web image data has become an increasingly important area of research in recent years. However, from related work, it is reflected that several gaps in the existing literature still need to be addressed. The authors highlight some of the key gaps in the literature related to the analysis of dark web image data.

- **Limited scope:** Most existing methods and techniques that consider analyzing dark web image data are focused on a particular domain, such as Child Sex Abuse (CSA) or Weapons. This limited scope means that much dark web image data related to other categories is left out.
- **Incomplete coverage:** Even when some research considers content from more than one class, the focus is often on some specific classes, leaving other classes out of consideration. For instance, TOR Image Categories (TOIC) and Darknet Usage Service Images (DUSI) consider some categories while leaving others out.
- **Lack of publicly available data:** The datasets used in these research are not public, which limits the ability of other researchers to build upon previous work. This lack of publicly available data is a significant barrier to developing more comprehensive and inclusive approaches to analyzing dark web image data.
- **Lack of correlation analysis:** The analysis of text data related to respective image data to confirm the correlation is not much done. This correlation between text and Image data can help accurately identify the type of hidden service. However, this lack of correlation analysis can limit the ability to identify the type of hidden service accurately.

3. Proposed Method

To address the gaps in the existing literature related to the analysis of dark web image data, the authors propose a methodology that involves a combination of techniques. This methodology includes dataset preparation, training the CNN model, and applying quantum transfer learning. Further, Tf-Idf is applied to confirm the type of hidden services identified with the image label. First, the authors labeled image classes using the CNN model. Then they leveraged the work of the CNN model using quantum transfer learning to improve accuracy and enable the analysis of previously unexplored dark web image data classes. Finally, after quantum transfer learning, the authors label classes of the textual data using Tf-Idf to confirm the image classification model's accuracy further. The following subsections detail each step of the proposed methodology.

3.1. Dataset Overview

The dataset is created by combining images from two sources.

Source 1: Custom crawled hidden service dataset: The dataset for the research consists of a set of crawled hidden services. The designed crawler collected different hidden Services [10]. The dataset had 50000 dark web hidden service pages weighing 12.2 GB. Each collected hidden service had a folder comprising several HTML, CSS, JavaScript, Image, and other files. Figure 1 depicts file types of crawled hidden service.

Dataset cleaning was performed in preparation for analysis. First, the hidden services folders containing images were separated from the rest of the dataset in data set creation. Image files could be identified by analyzing the file paths. Further, it was necessary to fetch images from hidden services folders once those folders were identified. Therefore, python libraries were used to separate hidden services folders with images from the dataset. The dataset contains 2884 images after the image filtering process.

The next step involved removing any irrelevant images after fetching the images. For example, there were duplicates, images of poor quality, or images irrelevant to the analysis. Manual inspection by human experts conducted the image exclusion. As a result, 76 images were rejected from the dataset as deemed irrelevant. As a result, the refined dataset contained 2808 images relevant for analysis. Still, not all images belonged to categories useful for dark web service labeling. For example, images with the 'Food' label were irrelevant for dark web service classification. Thus, four classes were identified to label dark web services: Cards, Devices, Hackers, and Money.



Fig. 1. Overview of crawled hidden service

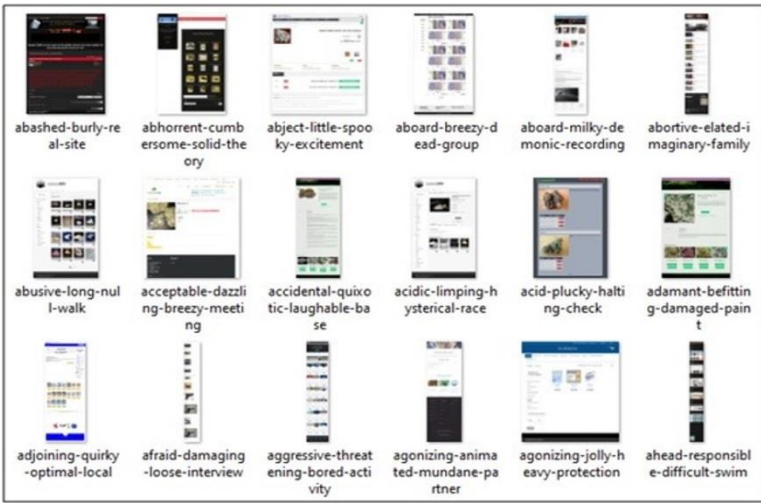


Fig. 2. CIRCL (Computer Incident Response Center Luxembourg) data source [10]

The dataset size of images belonging to the mentioned classes in crawled dark web services was 1090. The dataset size attempted to be further enhanced with an existing dataset of dark web snapshots.

Source 2: Dark Web Snapshots: CIRCL(Computer Incident Response Center Luxembourg)[1] is an open-source dataset comprising 8000 dark web hidden services snapshots, as in figure 2. To add images to the crawled hidden services dataset, images in dataset CIRCL were selected, mapping to Card, Device, Hacker, and Money classes. The newly created image dataset was named “Darkset”. Darkset has manually labeled 1650+ dark web images into four categories: Card, Device, Hacker, and Money.

The number of images per class is shown in Table 2.

Table 2. Number of Images per Class

Sr no.	Class Name	Number of images
1	Card	303
2	Device	958
3	Hacker	143
4	Money	295

3.2. Creating CNN Model

The authors used the Darkset dataset to train the CNN for the next step of the proposed methodology. The CNN model is designed as a sequential model that starts with a re-scale layer to standardize the input images. Re-scale layer enhances the image data and makes it more suitable for further processing. Next, four convolution layers follow the re-scale layer with an activation function called ReLU to help the model learn more complex representations of the input data.

Each convolution layer is followed by a max-pooling layer, which reduces the output data’s spatial dimensions and helps improve the model’s efficiency. After the convolution and pooling layers, a flatten layer converts the output data into a one-dimensional vector. Next, this vector is fed into dense, fully connected layers that perform nonlinear transformations on the input data. The last dense layer has classes equal to four, corresponding to the number of image categories in the dataset.

For the compilation of the model, the authors used the Adam optimizer. Adam is a gradient-based optimization algorithm that is used to update the parameters of the model during training. It combines the best features of two other optimization algorithms, stochastic gradient descent (SGD) and Adagrad, to provide better performance and faster convergence. The authors trained the model for a total of 30 epochs. An epoch is defined as one complete pass through the entire training dataset. Using multiple epochs, the model can learn more complex patterns in the data and improve its accuracy over time. The implemented algorithm to label the hidden service with is represented below as algorithm 1.

Algorithm 1 Image Class Prediction with CNN

Require: Input image x

Ensure: Output class probabilities y

1: Rescale x	▷ Rescale
layer	
2: Conv2D(32, (3,3), activation='relu')	▷ Convolution
layer 1	
3: MaxPooling2D((2,2))	▷Max-pooling
layer 1	
4: Conv2D(64, (3,3), activation='relu')	▷Convolution
layer 2	
5: MaxPooling2D((2,2))	▷Max-pooling
layer 2	
6: Conv2D(128, (3,3), activation='relu')	▷Convolution
layer 3	
7: MaxPooling2D((2,2))	▷Max-pooling
layer 3	
8: Conv2D(256, (3,3), activation='relu')	▷Convolution
layer 4	
9: MaxPooling2D((2,2))	▷Max-pooling
layer 4	
10: Flatten()	▷Flatten
layer	
11: Dense(128, activation='relu')	▷Dense
layer 1	
12: Dense(64, activation='relu')	▷Dense
layer 2	
13: Dense(4, activation='softmax')	▷ Dense layer with number of
classes = 4	▷ Compilation with adam
14: Compile(loss='categoricalCrossentropy', optimizer='adam')	▷ Training for 30
optimizer	
15: Fit(x , y , epochs=30)	
epochs	
16: $y \leftarrow \text{Predict}(x)$	▷ Prediction of output class
probabilities	
17: return y	

During the training process, training and validation datasets were used. Validation datasets were used to evaluate the model's performance during training, and training datasets were used to train the model. The proposed approach used a ratio of 80:20 for the training and validation dataset to prevent over-fitting by optimizing the model's performance.

3.3. Evaluating the performance of the CNN model

In this section, the authors discuss the accuracy and loss graphs to evaluate the CNN model, as shown in 3 and 4. The results showed that the CNN model could accurately classify images from the dark web dataset into their respective categories. Furthermore, the combination of convolutional, pooling, and dense layers in the model allowed it to learn complex representations of the input data and achieve high accuracy. To enhance the CNN model for images with multiple objects or regions of interest and images with occlusions or missing information, the authors propose leveraging quantum transfer learning methodology, which is discussed in the following section.

3.4. Quantum Transfer Learning

The authors propose a novel approach to extend the concept of transfer learning to hybrid classical-quantum neural networks to enhance the CNN model for image classification. The proposed methodology builds upon the general structure of transfer learning but with a crucial difference of using a quantum circuit to perform the final classification task. Figure 5 depicts the flow of quantum transfer learning. The following subsection presents a discussion on transfer learning.

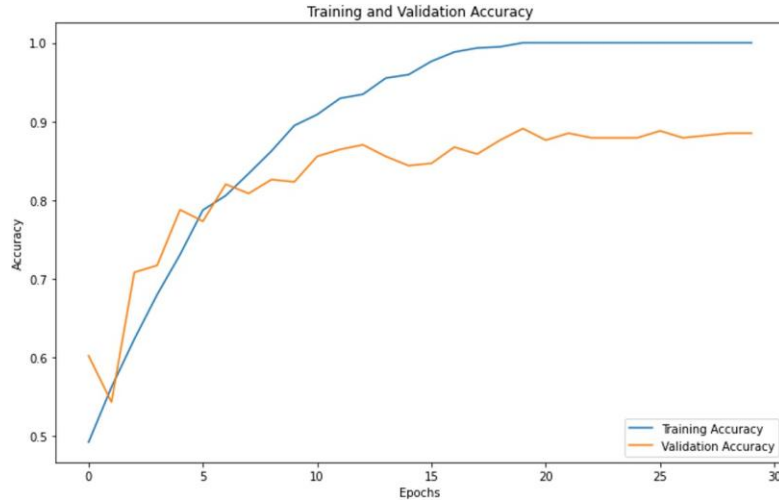


Fig. 3. Accuracy graph for training and validation set given to CNN model

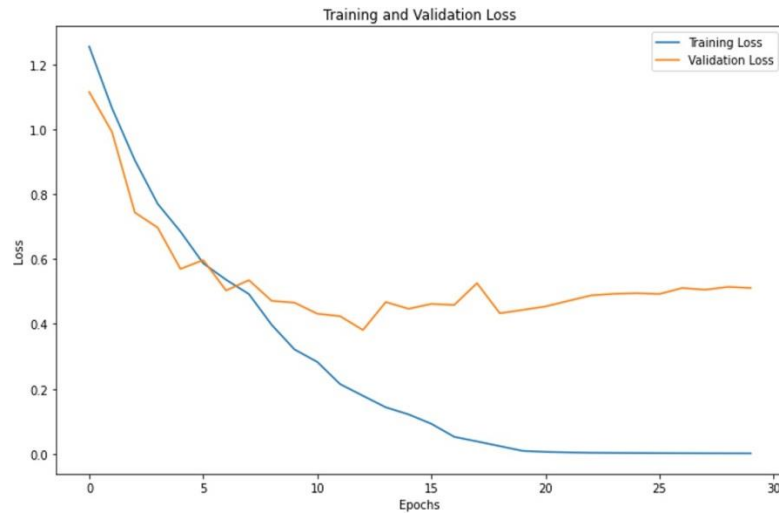


Fig. 4. Loss graph for training and validation set given to CNN model



Fig. 5. Flow of quantum transfer learning

3.4.1 Selection of Pre-trained network

The authors utilized the ResNet18 architecture as the pre-trained network due to its exceptional performance in various computer vision tasks, including object recognition, detection, and segmentation. ResNet18 is a deep residual network introduced in [17]. The ResNet18 architecture has 18 layers and is smaller than other state-of-the-art deep learning models, making it computationally efficient. In addition, its ability to extract high-level features from the input images makes it an ideal choice for the transfer learning task.

3.4.2 Defining the structure of the variational quantum circuit

Variational quantum circuits have been suggested as a potential quantum generalization of feedforward neural networks [11, 21, 23, 15, 29, 31, 32]. Like the classical case, a quantum layer can be defined as a unitary operation that acts on the input state of n_q quantum subsystems and generates the output state. A low-depth variational circuit can physically realize this quantum layer.

$$\zeta: |x\rangle \rightarrow y = U(w) |x\rangle \quad (1)$$

The array of classical variational parameters is denoted by w . The array of classical variational parameters is denoted by w . Quantum layers can take on various forms, such as a sequence of single-qubit rotations followed by a fixed sequence of entangling gates [31, 32], or a combination of active and passive Gaussian operations followed by single-mode non-Gaussian gates for optical modes [21]. It's important to note that, unlike classical layers, quantum layers maintain the Hilbert-space dimension of the input states. This is a result of the fundamental unitary nature of quantum mechanics, and should be considered when designing quantum networks.

To incorporate classical data into a quantum network, it is necessary to encode a real vector x into a quantum state represented as x .

A variational embedding layer can also be utilized for this purpose, which relies on x and is applied to a reference state like the vacuum or ground state.

$$\xi: |x\rangle \rightarrow x = E(x) |0\rangle \quad (2)$$

The embedding layer E is a map from a classical vector space to a quantum Hilbert space. The following paragraph discusses the layers used to create the variational quantum circuits.

$$M|x\rangle \rightarrow \langle x | \{y\} |x\rangle \quad (3)$$

Embedding Layer: The first step in a quantum algorithm is the embedding layer. At this stage, all qubits are initially prepared in a state of balanced superposition between the up and down states. This allows the qubits to be manipulated in a way that can represent and process the input data. Following initialization, the qubits are rotated based on the input parameters using a local embedding technique. This process is critical for the success of the quantum machine learning algorithm as it allows the input data to be encoded into the state of the qubits.

Variational Layers: After the embedding layer, a sequence of variational layers is applied in a quantum algorithm. The variational layers consist of trainable rotation and constant entangling layers, enabling the quantum computer to compute the encoded data. The rotation layers modify the state of the qubits to solve the problem at hand, while the entangling layers enable the qubits to become correlated.

Measurement Layer: The final step in a quantum algorithm is the measurement layer. At this stage, the local expectation value of the Z operator is measured for each qubit. The Z operator measures the spin of a qubit in the z direction. This measurement produces a classical output vector, which can be further processed to obtain the final result. The output vector is suitable for additional post-processing, such as classical optimization techniques or machine learning algorithms. This measurement step is critical as it enables the quantum computer to output a result that can be compared to the expected output of the problem.

In general, the complete quantum network, which involves both the initial embedding layer and the final measurement, can be expressed as:

$$Q = M \circ Q \circ \varepsilon \quad (4)$$

Algorithm 2 provides the algorithm to define the quantum layers.

Algorithm 2 Defining quantum layers in quantum circuit

```

1: function HLayer(nqubits)                                ▷ Layer of single-qubit Hadamard gates
2:     for idx ← 0 to nqubits – 1 do
3:         qml.Hadamard(wires = idx)
4:     end for
5: end function

6: function Ry_Layer(w)                                     ▷ Layer of parametrized rotations around the y-axis
7:     for IDX ← 0 TO length(w) – 1 do
8:         qml.RY (w[idx], wires = idx)
9:     end for
10: end function

11: function EntanglingLayer(nqubits)                       ▷ Layer of CNOTs followed by another shifted layer of CNOT
12:     for i ← 0 to nqubits – 2 by 2 DO                     ▷ Loop over even indices: i = 0, 2, ..., nqubits – 2
13:         qml.CNOT (wires = [i, i + 1])
14:     end for
15:     for i ← 1 to nqubits – 2 by 2 DO                     ▷ Loop over odd indices: i = 1, 3, ..., nqubits – 3
16:         qml.CNOT (wires = [i, i + 1])
17:     end for
18: end function

```

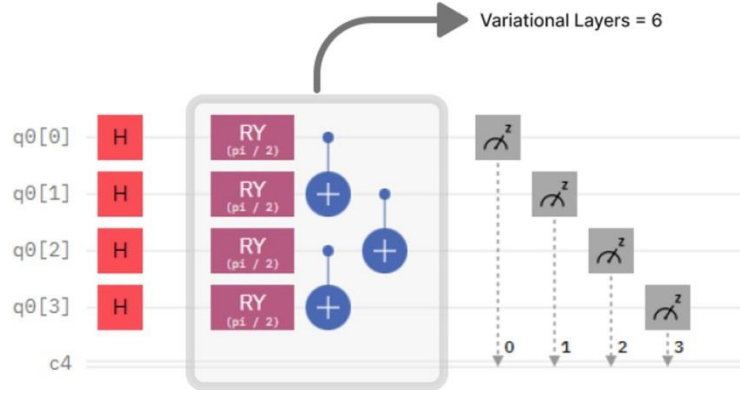


Fig. 6. Six variational layers quantum circuit

The authors used the quantum layers to design a custom quantum network for image classification tasks by replacing the fully connected final layer of the ResNet18 pre-trained model with a 4-qubit quantum layer. The quantum layer takes 512 features as input and performs the classification task, outputting four features. The quantum network consists of a classical preprocessing layer, a classical activation function, a constant scaling, the quantum circuit, and a classical postprocessing layer. The quantum circuit has a depth of 6 variational layers and was evaluated using both local simulation using intel i7 11th generation processor and NVIDIA RTX 3050 Ti and on IBM quantum computer with 1024 shots per epoch. The circuit outputs the Pauli Z expectation values. Figure 6 represents the four qubit circuits and six variational layers used.

The quantum net defining algorithm as defined in algorithm 3, describes the formulation of 6 layers variational circuit.

The authors further discuss designing dressed quantum net using variational circuits.

3.4.3 Dressed quantum net for image classification

In this section, the authors present their custom quantum network for dark web image classification, which utilizes transfer learning to leverage the benefits of both classical and quantum computing. To enable transfer learning across the classical-quantum interface, it is necessary to establish a connection between classical neural networks and quantum variational circuits. Given that classical and quantum networks may vary significantly in size, a more flexible quantum circuit model is preferable. Specifically, we consider a variational circuit based on n_q subsystems, as defined in Equation

(4). To incorporate basic pre-processing and post-processing of the input and output data, a classical layer is introduced at the outset and culmination of the quantum network. This arrangement yields a dressed quantum circuit, which offers enhanced versatility and adaptability for transfer learning purposes.

Algorithm 3 Defining quantum net

```

1: function QuantumNet(input_features, qWeights)                                ▷ Layer of single-qubit Hadamard gates
2:     Reshape weights
3:     qWeights ← qWeightsFlat.reshape(qDepth, nqubits)
4:     Start from state  $|+\rangle$ , unbiased w.r.t  $|0\rangle$  and  $|1\rangle$ 
5:     call function HLAYER(nqubits)

6:     call function Ry Layer(input_features)                                    ▷ Embed features in quantum node
7:     for k ← 0 TO qDepth - 1 do                                              ▷ Entangling layer (nqubits)
8:         call function Ry Layer(qWeights[k])
9:     end for
10:    Expectation ← [expect(PauliZ(position)) for position(nqubits)]
11:    Return tuple(Expectation)
18: end function

```

The variational circuit specified in Equation (4), which is based on n_q subsystems. To facilitate the addition of rudimentary pre-processing and post-processing of the input and output data, authors introduced a classical layer at the beginning and end of the quantum network, respectively. This augmentation produces a dressed quantum circuit, denoted by Q , which is expressed as a composite function of the bare quantum circuit Q and the classical layers $L_{n_{in} \rightarrow n_q}$ and $L_{n_q \rightarrow n_{out}}$. Here, $L_{n_{in} \rightarrow n_0}$ is derived from Equation (5)

$$L_{x \rightarrow 0}: x \rightarrow y = \varphi(Wx + b) \quad (5)$$

Authors designed the network which takes a batch of vectors with 512 real parameters from the pre-trained ResNet18 model and outputs two real outputs representing the two classes of images. The quantum circuit is applied to each element of the batch, and the resulting outputs are concatenated and passed through the quantum processing layer to produce the final prediction. The dressed quantum circuit is attached as the final layer of ResNet18, and the authors freeze the weights of the classical network while training only the quantum layer using the cross-entropy loss function and Adam optimizer. A learning rate scheduler reduces the learning rate by a gamma factor after every ten epochs. This approach achieves state-of-the-art results on image classification tasks. The algorithm is defined in algorithm 4.

Algorithm 4 Dressed Quantum Net

```

1: function DRESSEDQUANTUMNET
2:     Initialize pre-processing layer as a linear layer with 512 input neurons and nqubits output neurons
3:     Initialize quantum parameters as a tensor with qdepth * nqubits dimensions
4:     Initialize post-processing layer as a linear layer with n qubits input neurons and 2 output neurons
5: end function
6: function FORWARD(input features)
7:     Pass input features to pre-processing layer
8:     Compute the activation of the pre-processing layer using tanh activation function and scale by  $\pi/2$ 
9:     Apply the quantum circuit to each element of the batch and append the output to qout
10:    Return the two-dimensional prediction from the post-processing layer
11: end function
  
```

3.4.4 Evaluating the performance of the Quantum transfer learning model

The classical-quantum transfer learning model uses epochs to determine which model represents the sample with the lowest error. Before training the dressed neural network, the epoch and batch size must be specified. Figure 7 shows how the negative log-likelihood loss varies over training iterations.

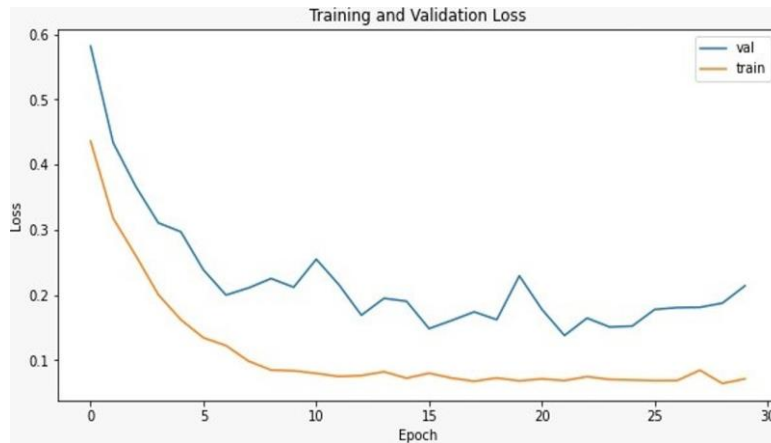


Fig. 7. Negative log-likelihood loss v/s training iterations

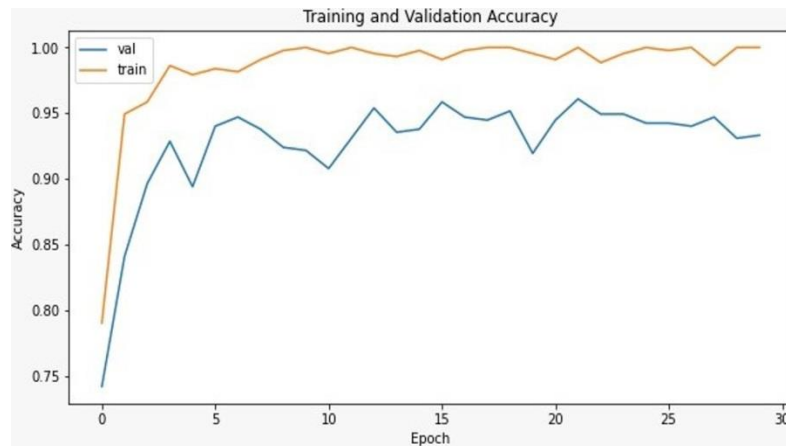


Fig. 8. Training and validation loss graph of quantum transfer learning model

Figure 8 shows the model loss graph; it shows the loss of the model changes throughout training. The x-axis represents the number of epochs, while the y-axis represents the loss of the model on the validation dataset. The loss represents how well the model can predict the correct output for a given input. The graph determines if the model is improving during training and if any changes need to be made to the model architecture or hyperparameters. It can also be used to compare the performance of different models; the proposed classical-quantum transfer learning model has the lowest loss and is considered the best-performing model.

Figure 9 shows the model accuracy graph, the accuracy of the model changes throughout training. The x-axis represents the number of epochs or completed passes through the training dataset, while the y-axis represents the model's accuracy on the validation dataset. The graph is used to visualize the model's progress during training and to determine if the model is fitting or underfitting the data.

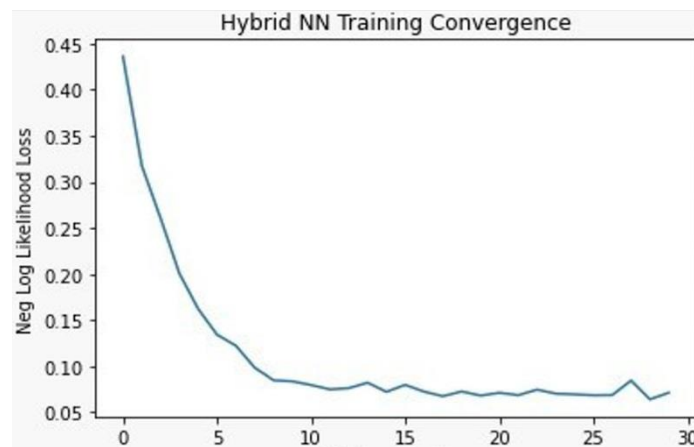


Fig. 9. Training and Validation accuracy graph of quantum transfer learning model

The proposed classical-quantum transfer learning model is well-performing with high accuracy on given datasets, with minimal difference between them.

3.5. Comparison of CNN model with classical-quantum transfer learning

The proposed methodology combines Convolutional Neural Networks (CNN) for image classification and Quantum Transfer Learning to enhance the learning process, providing a robust system for labeling images from different content categories. The CNN-based model performs well in identifying various classes of hidden services, including device and hacker classes. However, further analysis demonstrates that quantum transfer learning outperforms the CNN model's accuracy.

The results show that the proposed method could improve image recognition performance by leveraging the power of quantum computing with transfer learning techniques. Specifically, using quantum-enhanced classifiers obtained through transferring knowledge between models trained on classical data sets to models tailored to separate hidden service images leads to remarkable improvements over baseline classification models.

This study offers compelling evidence that Quantum Transfer Learning is a promising approach for improving image recognition tasks concerning dark web content classification beyond what traditional deep neural networks offer. Furthermore, it exemplifies how combining cutting-edge technologies from multiple domains, such as Deep Learning, Quantum computing, and information retrieval, could be beneficial and create breakthroughs in domain-specific problems' techno- logical frontiers.

CNNs have shown promising results in analyzing static imagery from online environments such as anonymity networks like Tor, and this research indicates that they may not always adequately analyze digital content generated within darknets.

3.6. Correlating hidden services text with images

The further discussion outlines the text analysis process for the proposed approach that examined scraped data from the hidden services collected from the dark web. Analyzing dark web hidden services text involves the term frequencyinverse document frequency (TF-IDF) feature engineering approach. Text analysis through TF-IDF has proven useful for sentiment analysis and topic modeling tasks dealing with sizeable dark web datasets containing numerous types of textual information. As such, this research methodology contributes significantly to uncovering potentially insightful perspectives about dark web services.

TF-IDF is a widely used algorithmic technique in natural language processing that helps evaluate word importance within a document. It calculates term frequency (TF) and inverse document frequency (IDF). Term frequency refers to how often a given word appears in a particular document, while inverse document frequency measures its significance relative to other documents in the dataset. The combined score obtained using TF-IDF provides valuable insights into which words are most important in context - helping researchers identify topics or themes that may exist across multiple documents.

To carry out this analysis, the author utilized BeautifulSoup as an HTML parser and a TF-IDF feature extraction pipeline. This process involved taking plain-text information scraps collected from various sources on the dark web services and further classifying them into image class topics based on significant terms mentioned in the text content of the hidden services. Figure 10 shows the flow of dark web text analysis using the TF-IDF approach.

4. Result and Discussion

In this section, the authors present the results and discussion of this study, which aimed to classify images and textual data from dark web hidden services. The following subsection discusses the image and textual prediction based on the proposed models.

4.1. Image class prediction

A sample of random images was tested to evaluate model performance. Table 3 contains data that shows predictions of randomly selected 16 sampled images using CNN and quantum transfer learning models.

In Table 3 in the Prediction by CNN model and prediction by quantum transfer learning model signifies “incorrect prediction,” whereas 1 signifies “correct prediction”. So, for example, a class of 11 Images out of 16 is predicted correctly by the CNN model, and the quantum transfer learning model correctly predicts a class of 15 images out of 16.

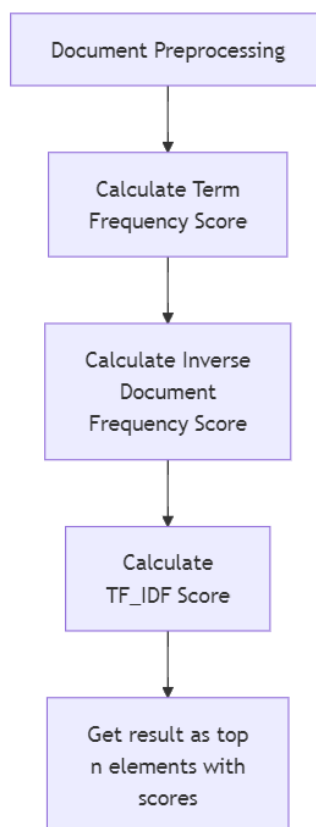


Fig. 10. Flow of dark web text analysis using the TF-IDF approach

Table 3. Image predictions by CNN and quantum transfer learning

Sr No.	Image	Class of Image	Class Pre- dicted by CNN	Prediction by CNN	Class Pre- dicted by QTL	Prediction by QTL
1	Image1	Device	Device	1	Device	1
2	Image2	Money	Device	0	Money	1
3	Image3	Hacker	Hacker	1	Hacker	1
4	Image4	Device	Device	0	Device	1
5	Image5	Hacker	Device	0	Hacker	1
6	Image6	Hacker	Hacker	1	Hacker	1
7	Image7	Device	Device	1	Device	1
8	Image8	Device	Device	1	Device	1
9	Image9	Card	Money	0	Card	1
10	Image10	Card	Card	1	Device	0
11	Image11	Card	Money	0	Card	1
12	Image12	Hacker	Hacker	1	Hacker	1
13	Image13	Money	Money	1	Money	1
14	Image14	Money	Money	1	Money	1
15	Image15	Money	Money	1	Money	1
16	Image16	Card	Hacker	0	Card	1

4.2. Hidden service labeling

The results of two randomly selected dark web hidden services are analyzed. Predicted labels with related text features obtained using TF-IDF are demonstrated in tables 4, 5, 6, and 7. Analysis of image and text data for any hidden service helps understand the topic the hidden service is about or the type of hidden service.

Sr No	Image	Class of Image	Class Predicted	0=Predicted Wrongly 1=Predicted Rightly
1	3ED329ImEA3CD420YA5o9F8BE3	Hacker	Hacker	1
2	9719E67F25IEF8BFAF5AD516C1rB	Device	Hacker	0
3	D5CCB75Bm237B12281A0BFB1VCEA8	Hacker	Hacker	1

Sr No	HTML File	TF-IDF Output
1	0A192H9622eBDfE28401186Wqe	2020_reply,says,facebook,hack
2	0EBEAE13GF3CC9506F6tB15EC95	0,eghix,contexthttpsschemaorggraphypewebseidhttpswwwwinstabitn etworkcomwebsiteurlhttpswwwwinstabitnetworkcomnamehire,onlined escriptionhackers,servicepotentialactiontypeasearchactiontargethttps wwwinstabitnetworkcomssearchtermstringqueryinputrequired
3	1F85YCRkF20B4BCj41F3A207	10hire,onlinehttpswwwwinstabitnetworkcomhire,onlinehttpswwwwinsta bitnetworkcomauthorarchitectverified,real,hackermich600338blockqu ete

Fig. 11. Identification of sample hidden service 1 with image class and text feature analysis

Sr No	Image	Class of Image	Class Predicted	0=Predicted Wrongly 1=Predicted Rightly
1	_273E7DE0AC15AF823Ay5EA3DA9F	Hacker	Hacker	1
2	3DB6D1zSsepBCE325perF204F893WA5	Device	Device	1

Sr No	HTML File	TF-IDF Output
1	0EY1501EAE2D683B492EAcB0E8BCEB	1,2020,western,i've,hacking
2	5D3D-\$1598169L10B7F00U9D	Var,blank,hacking,function,contexthttpsschemaorggraphyp eorganizationidhttpswwesternunionbugsorganizationname western
3	8C5B09293EFaxbF5C4IC65C9F3u	Object,objectypeintegercontextextrequiredfalsedefaultviewe numviewembededitdescriptionscope.password.iso8601.pu blished
4	25F306DE2BF8111E323DB1C6853BJ_k2dubT9	Function,hacking,hackers,var,noopfn

Fig. 12. Identification of sample hidden service 2 with image class and text feature analysis

5. Limitation

While authors approach offers valuable insights into dark web hidden services through image classification and quantum transfer learning, several limitations should be noted.

First, the manual classification of images into just four categories—Cards, Devices, Hackers, and Money—may not capture the full spectrum of illegal activities on the dark web. This limited scope could result in significant hidden services being overlooked. Additionally, relying on a curated dataset introduces biases that may affect how well the model performs on unseen data.

Moreover, our methodology leans heavily on image data, which might not always represent the true nature of hidden services. The impact of associated text data on classification isn't fully explored, meaning text-heavy services

might not be accurately identified.

On the quantum computing side, while Quantum Transfer Learning (QTL) shows promise, it comes with challenges. Controlling entangled qubits is complex, and issues like quantum decoherence can affect the model's performance. Even the best datasets may not yield optimal results if the quantum system isn't stable. Current quantum hardware also has limitations in the number of qubits that can be effectively managed, which can restrict the advantages of using quantum methods.

Lastly, our dark web crawler might not capture all hidden services due to the constantly evolving nature of the dark web. Continuous updates will be necessary to keep up with new content and emerging services.

6. Conclusion

The proposed work introduces a novel approach to labeling hidden services by correlating the image and text of the particular hidden service. Convolutional Neural Networks (CNN) models are leveraged for image classification, and Quantum Transfer Learning (QTL) is utilized to enhance the learning process. The implementation of QTL further enhances learning. Specifically, a pre-trained ResNet18 network is selected for image classification tasks. In integrating QTL, the final classification and convolution layers of ResNet18 are removed and replaced with a Quantum Neural Network (QNN), also known as a quantum net, constructed using a quantum variational circuit to incorporate quantum principles into the learning process.

Additionally, TF-IDF analysis is employed for textual data associated with hidden services. The study contributes to labeling images from four new dark web hidden service categories - Cards, Devices, Hackers, and Money. Moreover, the correlation between text and image analysis is established through confirmatory evidence-based results using semantic similarity between identified important words from individual text files and labeled images of dark web hidden services.

The conclusion of the work affirms that categorizing dark web content using images increases complexity but also enhances accuracy, especially in cases where text information is limited or unavailable. Image analysis provides additional details to aid categorization when text information is lacking. For Law Enforcement Agencies (LEAs), identifying and blocking inappropriate content on web pages is facilitated by analyzing images. Hence, it is concluded that although image classification poses complexity, it ultimately improves the overall effectiveness of categorizing and filtering dark web content.

References

- [1] Computer Incident Response Center Luxembourg (CIRCL). CIRCL AIL Dataset 01. <https://www.circl.lu/opendata/datasets/circl-ail-dataset-01>. Accessed: April 3, 2023. 2018.
- [2] Mhd Wesam Al Nabki et al. "Classifying illegal activities on tor network based on web textual contents". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 2017, pp. 35–43.
- [3] Randa Basheer and Bassel Alkhatib. "Threats from the dark: a review over dark web investigation research for cyber threat intelligence". In: *Journal of Computer Networks and Communications* 2021 (2021), pp. 1–21.
- [4] Rubel Biswas, Eduardo Fidalgo, and Enrique Alegre. "Recognition of service domains on TOR dark net using perceptual hashing and image classification techniques". In: *8th International Conference on Imaging for Crime Detection and Prevention (ICDP 2017)*. IET. 2017, pp. 7–12.
- [5] Rubel Biswas et al. "Perceptual Hashing applied to Tor domains recognition". In: *arXiv preprint arXiv:2005.10090* (2020).
- [6] Gwern Branwen et al. *StExo, El Presidente, Anonymous, Daryl Lau, Delyan Kratunov Sohhlz, Vince Cakic, Van Buskirk, Whom, Michael McKenna, and Sigi Goode. 2015. Dark net market archives, 2011-2015*. 2021.
- [7] Madeleine van der Bruggen and Arjan Blokland. "Child sexual exploitation communities on the Darkweb: How organized are they?" In: *Cybercrime in Context: The human factor in victimization, offending, and policing*. Springer, 2021, pp. 259–280.
- [8] Janis Dalins, Campbell Wilson, and Mark Carman. "Criminal motivation on the dark web: A categorisation model for law enforcement". In: *Digital Investigation* 24 (2018), pp. 62–71.
- [9] Ashwini Dalvi et al. "Content Labelling of Hidden Services With Keyword Extraction Using the Graph Decomposition Method". In: *Using Computational Intelligence for the Dark Web and Illicit Behavior Detection*. IGI Global, 2022, pp. 181–205.
- [10] Ashwini Dalvi et al. "SpyDark: Surface and Dark Web Crawler". In: *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*. IEEE. 2021, pp. 45–49.
- [11] Edward Farhi and Hartmut Neven. "Classification with quantum neural networks on near term processors". In: *arXiv preprint arXiv:1802.06002* (2018).
- [12] Eduardo Fidalgo Fernandez et al. "Classifying suspicious content in Tor Darknet". In: *arXiv preprint arXiv:2005.10086* (2020).
- [13] Eduardo Fidalgo et al. "Classifying suspicious content in tor darknet through Semantic Attention Keypoint Filtering". In: *Digital Investigation* 30 (2019), pp. 12–22.
- [14] Eduardo Fidalgo et al. "Illegal activity categorisation in DarkNet based on image classification using CREIC method". In: *International Joint Conference SOCO'17-CISIS'17-ICEUTE'17 León, Spain, September 6–8, 2017, Proceeding 12*. Springer. 2018, pp. 600–609.
- [15] Michael R Geller. "Rigorous measurement error correction". In: *Quantum Science and Technology* 5.3 (2020), 03LT01.
- [16] Mahdi Hashemi and Margeret Hall. "Detecting and classifying online dark visual propaganda". In: *Image and Vision Computing* 89 (2019), pp. 95–105.

- [17] 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.
- [18] Siyu He, Yongzhong He, and Mingzhe Li. “Classification of illegal activities on the dark web”. In: *Proceedings of the 2nd International Conference on Information Science and Systems*. 2019, pp. 73–78.
- [19] Susan Jeziorowski, Muhammad Ismail, and Ambareen Siraj. “Towards image-based dark vendor profiling: An analysis of image metadata and image hashing in dark web marketplaces”. In: *Proceedings of the Sixth International Workshop on Security and Privacy Analytics*. 2020, pp. 15–22.
- [20] Shubhdeep Kaur and Sukhchandan Randhawa. “Dark web: A web of crimes”. In: *Wireless Personal Communications* 112 (2020), pp. 2131–2158.
- [21] Nathan Killoran et al. “Continuous-variable quantum neural networks”. In: *Physical Review Research* 1.3 (2019), p. 033063.
- [22] Roberta Liggett et al. “The dark web as a platform for crime: An exploration of illicit drug, firearm, CSAM, and cybercrime markets”. In: *The Palgrave handbook of international cybercrime and cyberdeviance* (2020), pp. 91–116.
- [23] Jarrod R McClean et al. “The theory of variational hybrid quantum-classical algorithms”. In: *New Journal of Physics* 18.2 (2016), p. 023023.
- [24] Harsha Moraliyage et al. “Multimodal Classification of Onion Services for Proactive Cyber Threat Intelligence Using Explainable Deep Learning”. In: *IEEE Access* 10 (2022), pp. 56044–56056.
- [25] Saiba Nazah et al. “An unsupervised model for identifying and characterizing dark web forums”. In: *IEEE Access* 9 (2021), pp. 112871–112892.
- [26] Javier Pastor-Galindo, Fe’lix Go’mez Ma’rmol, and Gregorio Mart’inez Pe’rez. “On the gathering of Tor onion addresses”. In: *Future Generation Computer Systems* 145 (2023), pp. 12–26.
- [27] Stephan Raaijmakers. “Artificial intelligence for law enforcement: challenges and opportunities”. In: *IEEE security & privacy* 17.5 (2019), pp. 74–77.
- [28] Matthias Scha’fer et al. “BlackWidow: Monitoring the dark web for cyber security information”. In: *2019 11th International Conference on Cyber Conflict (CyCon)*. Vol. 900. IEEE. 2019, pp. 1–21.
- [29] Maria Schuld et al. “Circuit-centric quantum classifiers”. In: *Physical Review A* 101.3 (2020), p. 032308.
- [30] Vidyesh Shinde et al. “CrawlBot: A Domain-Specific Pseudonymous Crawler”. In: *Cybersecurity in Emerging Digital Era: First International Conference, ICCDE 2020, Greater Noida, India, October 9-10, 2020, Revised Selected Papers*. Springer. 2021, pp. 89–101.
- [31] Sukin Sim, Peter D Johnson, and Ala’n Aspuru-Guzik. “Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms”. In: *Advanced Quantum Technologies* 2.12 (2019), p. 1900070.
- [32] Sukin Sim, Peter D Johnson, and Ala’n Aspuru-Guzik. “Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms”. In: *Advanced Quantum Technologies* 2.12 (2019), p. 1900070.
- [33] EDA So’ñmez and Keziban Sec’kin Codal. “Terrorism in cyberspace: A critical review of dark web studies under the terrorism landscape”. In: *Sakarya University Journal of Computer and Information Sciences* 5 (2022).
- [34] Xiangwen Wang et al. “You are your photographs: Detecting multiple identities of vendors in the darknet marketplaces”. In: *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. 2018, pp. 431–442.
- [35] Ning Zhang et al. “Counteracting dark Web text-based CAPTCHA with generative adversarial learning for proactive cyber threat intelligence”. In: *ACM Transactions on Management Information Systems (TMIS)* 13.2 (2022), pp. 1–21.

Authors’ Profiles



Ashwini Dalvi received the M.Tech. Degree in Computer Engineering from the Veermata Jijabai Technological Institute, Mumbai, India, where she is currently pursuing a Ph.D. degree in Computer Engineering. She is an Assistant Professor at K J Somaiya College of Engineering. Her research interests include Dark web Monitoring, Cyber Security, and Intelligent applications.



Soham Bhoir is currently pursuing a BTech in Information Technology at K. J. Somaiya College of Engineering. His research interests revolve around Quantum Information Processing and Natural language processing, particularly exploring applications of quantum deep learning.



Akansha Singh has recently completed her MTech in Computer Engineering from Veermata Jijabai Technological Institute (VJTI), Mumbai, India. She is currently employed as a software developer at eClerx.



Irfan Siddavatam received his PhD in Electronics Engineering from Veermata Jijabai Technological Institute, Mumbai in 2018. He is a Professor at Somaiya School of Design and an Adjunct Professor in the Department of Information Technology at K.J.Somaiya College of Engineering, where he has been since 2001. His research interests include Information and Cybersecurity, Artificial Intelligence, Machine Learning, and Blockchain.



S. G. Bhirud is currently Vice Chancellor at the College of Engineering Pune (COEP) Technological University Pune. He has more than twenty years of professional experience in data mining, information systems, and technologies. He has published more than 50 research articles in leading journals and conference proceedings.

How to cite this paper: Ashwini Dalvi, Soham Bhoir, Akansha Singh, Irfan Siddavatam, Sunil Bhirud, "Identifying Dark Web Hidden Services with Novel Image Classes Using CNN and Quantum Transfer Learning", International Journal of Education and Management Engineering (IJEME), Vol.15, No.2, pp. 44-59, 2025. DOI:10.5815/ijeme.2025.02.05