IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT
Impact Factor: 6.252

Volume 2, Issue 6, June 2022

Face Authentication in Chat Application using Cryptography

Supriya B C[1], Thanuja K[2], Thanuja R[3], Prof. Prathibha R[4]

Final Year Student, Department of Information Science and Engineering[1,2,3]
Assistant Professor, Department of Information Science and Engineering[4]
S J C Institute of Technology, Chikkaballapur, Karnataka, India

Abstract: *Data plays a vital role in the world of computer network. There are many platforms to share data between a group of devices or individual or a group of peoples. Every person in the community needs him Information to keep safe from a third party. Here we have created a simple chat application with face authentication that can be used once to view end-to-end encryption authentication and shared message. This takes credentials to register and requires Face snap to authenticate the user. There are face snaps. Trained using the KNN Classifier. These face snaps are classified using the Haar cascade classifier algorithm for detection. User's face can be identified and logged in only if the user matches the credentials and can send messages. The receiving user is identified and the message is viewed.*

Keywords: Big four, Haar cascade classifier, Advance encryption standard, k-nearest neighbor, Authentication

I. INTRODUCTION

Sharing information or data in a secure manner plays a key role in the current environment. The information to be transferred over the network is encrypted at the sender's end and then sent via the wireless channel on the network, after which the information is decrypted at the end of the receiver and finally the message / information is received by the receiver. The encryption and decryption process can be done using the AES algorithm. When information is transferred through a channel, attackers are more likely to view the information and they can edit the message and send the message back to the receiver using the sender's name.

To prevent this, we use the Face Recognition technique, which is used on both sides when logging into the account and when the message arrives. It uses AES algorithm and facial recognition technology for secure communication between sender and receiver. Information plays a major role in a computer network. There are many different platforms for sharing information between groups of devices or individuals. Every person in the community needs to keep his information safe from a third party. Here we have created a simple chat application with biometric authentication, which gives end-to-end encryption and can be used once to view the shared message. It takes credentials to register and asks Face Snaps to authenticate the user. Face snaps are trained using the KNN Classifier. These face snaps are classified using the LBPH algorithm for detection. If the user recognizes his face and matches the user details and can send messages, the user will only be able to identify and log in when they view the message.

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

The name machine learning was coined in 1959 by Arthur Samuel. Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured

DOI: 10.48175/IJARSCT-5161

by P, improves with experience E." This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms.

Machine learning and data mining often employ the same methods and overlap significantly, but while machine learning focuses on prediction, based on known properties learned from the training data, data mining focuses on the discovery of (previously) unknown properties in the data (this is the analysis step of knowledge discovery in databases). Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve accuracy.

Much of the confusion between these two research communities (which do often have separate conferences and separate journals, ECML PKDD being a major exception comes from the basic assumptions they work with: in machine learning, performance is usually evaluated with respect to the ability to reproduce known knowledge, while in knowledge discovery and data mining (KDD) the key task is the discovery of previously unknown knowledge.

Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of some loss function on a training set of examples. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances (for example, in classification, one wants to assign a label to instances, and models are trained to correctly predict the pre-assigned labels of a set of examples). The difference between the two fields arises from the goal of generalization: while optimization algorithms can minimize the loss on a training set, machine learning is concerned with minimizing the loss on unseen samples.

II. RELATED WORK

2.1 Problem Statement

The Internet offers a cheap and convenient way to explore and communicate with distant people. A wide variety of services converge on the smartphone platform and the most important is social networking. With increased interconnectivity and the use of online services, concerns about consumer security and privacy are growing. Authentication is done using customer facial recognition, which is very reliable in identifying and identifying the right user.

2.2 Existing Method

- Privacy Protects consumer data from digital invaders when traveling over the Internet.
- Privacy is complex and expensive, especially for applications that involve communication and data exchange between multiple users.
- Cryptography is the most widely used medium to achieve privacy.
- It is a great challenge to share a secret key with a group of people in a secure and effective way.
- Secure communication between individuals ensures that the user is authenticated for each message received.
- Messages are no longer available to the user once the user has returned from the conversation portion.
- Authentication is done using customer facial recognition, which is very reliable in identifying and identifying the right user.
- Messages will not be displayed in the conversation even after they have been sent.
- Messages sent between users are stored in an encrypted format in the database, which enhances the security of the system.

III. METHODOLOGY

- To detect the face from the frame we are using HAAR Cascade classifier.
- To recognize the face we are using K-Nearest Neighbor (KNN) classification.
- To encrypt and decrypt the messages we are using Advanced Encryption Standard (AES).

3.1 HAAR Cascade Classifier

It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. The algorithm has four stages:

- Haar Feature Selection

- Creating Integral Images
- Adaboost Training
- Cascading Classifiers

It is well known for being able to detect faces and body parts in an image, but can be trained to identify almost any object. First step is to collect the Haar Features. A Haar feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums.

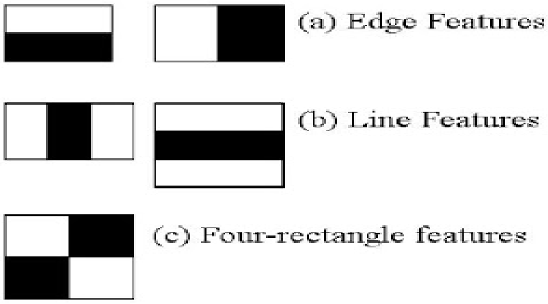


Fig 1: Features of Haar Cascade Classifier

But among all these features we calculated, most of them are irrelevant. For example, consider the image below. Top row shows two good features. The first feature selected seems to focus on the property that the region of the eyes is often darker than the region of the nose and cheeks. The second feature selected relies on the property that the eyes are darker than the bridge of the nose. But the same windows applying on cheeks or any other place is irrelevant.

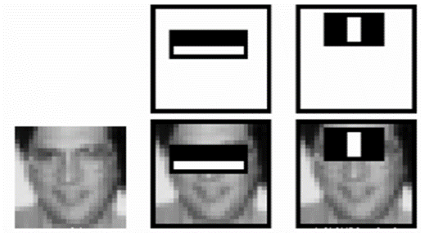


Fig 2: Face Features of Haar Cascade Classifier

So how do we select the best features out of 160000+ features? This is accomplished using a concept called Adaboost which both selects the best features and trains the classifiers that use them. This algorithm constructs a “strong” classifier as a linear combination of weighted simple “weak” classifiers. The process is as follows During the detection phase, a window of the target size is moved over the input image, and for each subsection of the image and Haar features are calculated. You can see this in action in the video below. This difference is then compared to a learned threshold that separates non-objects from objects. Because each Haar feature is only a "weak classifier" (its detection quality is slightly better than random guessing) a large number of Haar features are necessary to describe an object with sufficient accuracy and are therefore organized into ***cascade classifiers*** to form a strong classifier.

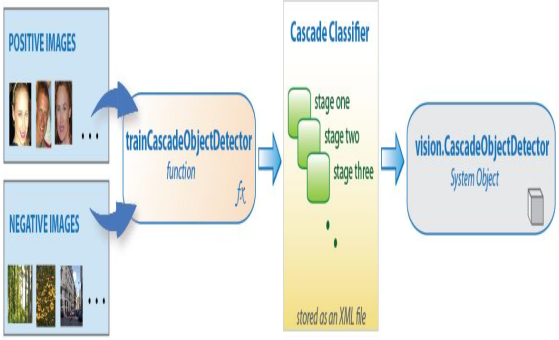
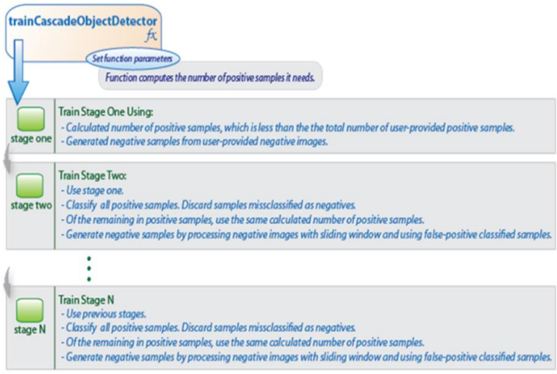


Fig 3: Model of Cascade Classifier

The cascade classifier consists of a collection of stages, where each stage is an ensemble of weak learners. The weak learners are simple classifiers called *decision stumps*. Each stage is trained using a technique called boosting. *Boosting* provides the ability to train a highly accurate classifier by taking a weighted average of the decisions made by the weak learners. Each stage of the classifier labels the region defined by the current location of the sliding window as either positive or negative.

Positive indicates that an object was found and *negative* indicates no objects were found. If the label is negative, the classification of this region is complete, and the detector slides the window to the next location. If the label is positive, the classifier passes the region to the next stage. The detector reports an object found at the current window location when the final stage classifies the region as positive.



Function computes the number of positive samples it needs.

Train Stage One Using:
- Calculated number of positive samples, which is less than the the total number of user-provided positive samples.
- Generated negative samples from user-provided negative images.

Train Stage Two:
- Use stage one.
- Classify all positive samples. Discard samples missclassified as negatives.
- Of the remaining in positive samples, use the same calculated number of positive samples.
- Generate negative samples by processing negative images with sliding window and using false-positive classified samples.

Train Stage N
- Use previous stages.
- Classify all positive samples. Discard samples missclassified as negatives.
- Of the remaining in positive samples, use the same calculated number of positive samples.
- Generate negative samples by processing negative images with sliding window and using false-positive classified samples.

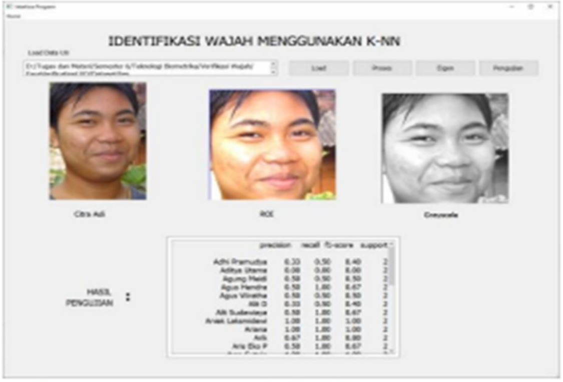
Fig 4: Stages of Cascade Classifier

A *false negative* occurs when a positive sample is mistakenly classified as negative. To work well, each stage in the cascade must have a low false negative rate. If a stage incorrectly labels an object as negative, the classification stops, and you cannot correct the mistake. However, each stage can have a high false positive rate. Even if the detector incorrectly labels a nonobject as positive, you can correct the mistake in subsequent stages. Adding more stages reduces the overall false positive rate, but it also reduces the overall true positive rate. Cascade classifier training requires a set of positive samples and a set of negative images. You must provide a set of positive images with regions of interest specified to be used as positive samples. You can use the Image Labeler to label objects of interest with bounding boxes. The Image Labeler outputs a table to use for positive samples. You also must provide a set of negative images from which the function generates negative samples automatically. To achieve acceptable detector accuracy, set the number of stages, feature type, and other function parameters.

3.2 KNN Classifier

Face classification is a stage for the process of matching testing data and training data from face datasets. KNN is one of the simple algorithms that can be The basic concept of KNN is to have several training samples and testing samples determined by members. If k=1, the testing sample is assigned to the nearest single neighbor class. However, finding the right k value for a particular problem is a problem that affects the performance of the KNN. The classification stages in face identification systems use the K-Nearest Neighbor (KNN) method where the eigen image of the feature extraction process used as input is as follows.

The KNN Classifier assumes that the objects that are similar in nature are exist within the closest distance. The Classifier is used to train the model based on the images provided by the user during the adding photos phase of the process. It trains the model using similarity of the photos of the user and classifies the images as belonging to the user. The algorithm for the same is given as “Initially the data is loaded in to the model, then the k(number of neighbors) is chosen. Next, for each data in the dataset, calculate the distance between the query points and the current example from the data and the distance is added with the index of example to an ordered collection. Sort the ordered collection of distances and indices in the ascending order of the distances. Pick the first of k entries from the sorted collection, get the label of the selected k entries. If regression then return the mean of the k labels else, return the mode of the labels”.



IDENTIFIKASI WAJAH MENGGUNAKAN K-NN

Load Data Uji

E:/Tugas dan Materi/Semester 6/Teknologi Biometrika/Verifikasi Wajah/

HASIL PENGUJIAN :

```
                    precision    recall  f1-score   support

     Adhi Pramudya       0.33      0.50      0.40         2
      Aditya Utama       0.00      0.00      0.00         2
       Agung Maldi       0.50      0.50      0.50         2
       Agus Hendra       0.50      1.00      0.67         2
      Agus Viratha       0.50      0.50      0.50         2
             Aik D       0.33      0.50      0.40         2
     Aik Sudewiaya       0.50      1.00      0.67         2
  Aniek Lakamidewi       1.00      1.00      1.00         2
            Ariana       1.00      1.00      1.00         2
              Arik       0.67      1.00      0.80         2
         Aria Eko P      0.50      1.00      0.67         2
```

Fig 5: Feature Extraction Process

IV. ALGORITHM

4.1 AES Algorithm

AES is an iterative rather than Feistel cipher. It is based on 'substitution–permutation network'. It comprises of a series of linked operations, some of which involve replacing inputs by specific outputs (substitutions) and others involve shuffling bits around (permutations).

The number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. Each of these rounds calculated from the original AES key. uses a different 128-bit round key, which is

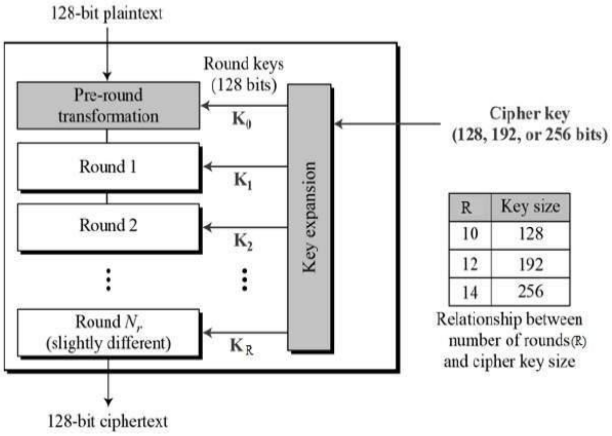


Fig 5: AES Algorithm

4.2 AES-Encryption Process

We restrict to description of a typical round of AES encryption. Each round comprise of four sub-processes. The first round process is depicted below

A. Byte Substitution (Sub Bytes)

The 16 input bytes are substituted by looking up a fixed table (S-box) given in design. The result is in a matrix of four rows and four columns.

B. Shift Rows

Each of the four rows of the matrix is shifted to the left. Any entries that 'fall off' are re-inserted on the right side of row. Shift is carried out as follows −

- First row is not shifted.

DOI: 10.48175/IJARSCT-5161

- Second row is shifted one (byte) position to the left.
- Third row is shifted two positions to the left.
- Fourth row is shifted three positions to the left.

The result is a new matrix consisting of the same 16 bytes but shifted with respect to each other.

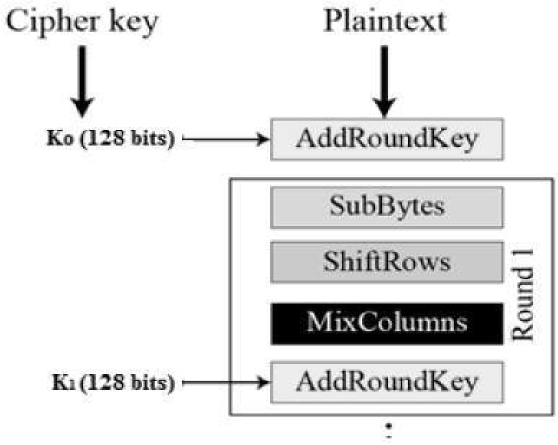


Fig 6: AES encryption Process

C. Mix Columns

Each column of four bytes is now transformed using a special mathematical function. This function takes as input the four bytes of one column and outputs four completely new bytes, which replace the original column. The result is another new matrix consisting of 16 new bytes. It should be noted that this step is not performed in the last round.

D. Add Round Key

The 16 bytes of the matrix are now considered as 128 bits and are XORed to the 128 bits of the round key. If this is the last round then the output is the ciphertext. Otherwise, the resulting 128 bits are interpreted as 16 bytes and we begin another similar round.

4.3 AES-Decryption Process

The process of decryption of an AES ciphertext is similar to the encryption process in the reverse order. Each round consists of the four processes conducted in the reverse order −

- Add round key
- Mix columns
- Shift rows
- Byte substitution

AES decryption algorithm is an iterated block decipher algorithm with a fixed block size of 128 and a variable key length. The AES algorithm operates on 128 bits of data and generates 128 bits of output. The length of the key used to decrypt this input data can be 128, 192 or 256 bits.

V. RESULTS

Despite all the limitations of facial recognition such as variations in posing, lighting and image quality, the technology is gaining popularity and will eventually become a part of consumers' daily lives. Platforms that provide simple implementation of facial recognition technology use a variety of algorithms and, therefore, can be used in a wide variety of applications. We try to show the experience gained in our specific project. As you can see, implementing a very simple idea of using facial recognition functionality is not so easy. We encountered some problems and tried to find solutions to work more effectively.

DOI: 10.48175/IJARSCT-5161

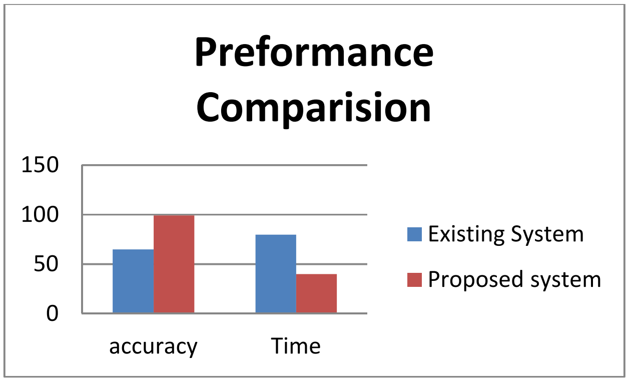


Fig 7: Result analysis

VI. CONCLUSION

The project, a simple chat application system with biometric authentication and encryption, was developed with the aim of securing communication between users. Biometric authentication is provided using the Haarcascade Classifier, KNN Classifier and Local Binary Patterns Histograms (LBPH) algorithm. Message encryption and decryption is done using the AES algorithm, which is a more secure method for the purpose of encryption and decryption available in the market. The encryption key used in the algorithm is 256 bits long. This increases system security, as the sender will not be able to see the sender's message until his face is authenticated. Also, messages are stored in the database in encrypted format so that messages stored in the database can also be viewed. The algorithms used for face authentication are very accurate in detecting the user's face even in low light and in different user face position.

VII. ACKNOWLEDGMENTS

We would like to express our gratitude to our College, SJC Institute of technology our Guide prof. Prathibha R mam and our project coordinator Aravind Tejas Chandra who have provided us with the opportunity to work on thisproject and given us support with guidance to make this project a success. We would also liketo thank our teammates for their contribution andcontinued support and zeal towards this project.This project wouldn't be a success without them.

REFERENCES

[1]. "Analyzing of Different Features Using Haar Cascade Classifier," Computerdoi: 10.1109/ICECOS.2018.8605266..
[2]. "Face Recognition Using Haar Cascade Classifier", ISSN:2349-5162, Vol.3,
[3]. "Face recognition system based on the improved LBPH," doi: 10.1109/ICCCIS48478.2019.8974493.
[4]. "Face recognition system based on the improved LBPH,"doi: 10.1109/ICCCIS48478.2019.8974493.
[5]. "A KNN Classifier for Face Recognition," 2021 International doi: 10.1109/CISCE52179.2021.9445908.
[6]. Selection: the applicant's point of view," pp. 1–11.
[7]. Wirdiani, Ayu & Hridayami, Praba & Widiari, Ayu & Rismawan, Komang & Candradinata, Putu & Jayantha, I. (2019). Face Identification Based on K-Nearest Neighbor. Scientific Journal of Informatics. 6. 150-159. 10.15294/sji.v6i2.19503.
[8]. N. Su, Y. Zhang and M. Li, "Research on Data Encryption Standard Based on AES Algorithm in Internet of Things Environment," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2019, pp. 2071-2075, doi: 10.1109/ITNEC.2019.8729488.
[9]. M Pitchaiah, Philemon Daniel, Praveen et al., "Implementation of Advanced Encryption Standard Algorithm," International Journal of Scientific and Engineering Research, ISSN: 2229-5518, Vol. 3, Issue 3, March-2012.
[10]. M. Khan, S. Chakraborty, R. Astya and S. Khepra, "Face Detection and Recognition Using OpenCV," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2019, pp. 116-119, doi: 10.1109/ICCCIS48478.2019.8974493.

DOI: 10.48175/IJARSCT-5161

[11]. S. Jayanthy, J. B. Anishkka, A. Deepthi and E. Janani, "Facial Recognition And Verification System For Accessing Patient Health Records," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1266-1271, doi: 10.1109/ICCS45141.2019.9065469.

[12]. Abhishek Pratap Singh, Sunil Kumar S Manvi,Pratik Nimbal, Gopalkrishna Shyam et al., “Face Recognition System based on LBPH algorithm,” International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249-8958, Vol.8, Issue 5S, May, 2019.

[13]. Varun Garg, Kritika Garg et al., "Face Recognition Using Haar Cascade Classifier", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.3, Issue 12, page no.140-142, December-2016.