International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

SKIN DISEASE PREDICTION USING ML

MUHAMMAD MAHSOOK S U [1] , SANJAY S [2], ABIRAMI N[3] , MOHAMMED AAKIF A J [4], SOMASUNDARAM A [5]

[1.]U.G. Student ,Department of computer science and application, Sri Krishna College of Arts and Science.
[2] U.G. Student ,Department of computer science and application, Sri Krishna College of Arts and Science.
[3]Assistant Professor ,Department of Computer Science, Sri Krishna College of Arts and Science, Coimbatore.
[4]U.G. Student ,Department of computer science and application, Sri Krishna College of Arts and Science.
[5] Assistant Professor ,Department of Computer Science, Sri Krishna College of Arts and Science, Coimbatore.
Bachelor of computer Science, Sri Krishna College of Arts and Science.

ABSTRACT :

Diagnosing skin diseases accurately and efficiently presents a significant challenge in healthcare. This project addresses this challenge by developing a machine learning-based system for skin disease prediction. The system analyzes digital images of skin conditions, extracts key features, and utilizes machine learning models to predict the probability of different skin diseases. This approach aims to reduce subjectivity in diagnosis, provide a supportive tool for dermatologists, and ultimately enhance patient care through improved diagnostic capabilities.

Keywords: Skin Disease Prediction, Machine Learning, Data Security, Privacy Protection, Medical Imaging

Introduction :

In the contemporary digital landscape, the security and integrity of data have emerged as paramount concerns. The rapid proliferation of data transmission over computer networks has led to a corresponding increase in the frequency and sophistication of network attacks. To mitigate these threats, it is crucial to encrypt data before it is transmitted or stored, thereby protecting it from various attackers and unauthorized access.

Encryption is a fundamental process in data security, involving the transformation of original data, often referred to as "plaintext," into an unreadable format known as "ciphertext."

This transformation is achieved through the application of encryption algorithms, which utilize sets of keys and characters to perform the encryption and subsequent decryption. During encryption, plaintext is converted into ciphertext, and the reverse process, decryption, involves converting the ciphertext back into the original plaintext.

Cryptography, the science dedicated to protecting data by encoding it into an unreadable form, plays a vital role in ensuring secure data transmission and storage. It is a crucial means of safeguarding sensitive information by employing mathematical algorithms for both the encryption and decryption processes. The effectiveness of an encryption algorithm is directly related to the difficulty involved in determining the key and recovering the original plaintext.

Encryption algorithms can be broadly categorized into two main types: symmetric and asymmetric, based on the keys used in the encryption and decryption processes. Symmetric algorithms utilize the same key for both encryption and decryption, and these algorithms are further divided into stream ciphers and block ciphers. Stream ciphers operate on individual bytes of data, while block ciphers operate on blocks of data. Asymmetric algorithms, on the other hand, employ two different keys: one for encryption and another for decryption. The security of the encrypted data depends on keeping the decryption key secret, thus preventing unauthorized decryption of messages.

Problem Definition :

The existing systems for image encryption face several limitations, primarily stemming from the unique characteristics of multimedia data, such as pixel correlation and high redundancy, which hinder the application of uniform protection techniquesTraditional encryption algorithms are not directly suited for images due to variations in size and structural differences compared to text, leading to potential inefficiencies in processing time and challenges in achieving exact equivalence between original and decrypted images. Furthermore, these systems often suffer from high computational demands, inefficiency in networking environments, and significant security vulnerabilities. To address these shortcomings, the proposed system aims to provide reliable storage and transmission of digital images, with a specific focus on enhancing image security through the implementation of the AES algorithm, which offers a secure encryption and decryption process, transforming images into a form that conceals their original content and ensuring robust protection against unauthorized access.

Proposed System :

The proposed system for skin disease prediction using machine learning aims to provide an automated and efficient solution for dermatological diagnosis. It will achieve this by employing a modular architecture that includes data acquisition, preprocessing, feature extraction, machine learning model development, and prediction/output components. The system may also include a security module to ensure the secure handling of patient data. This system will leverage technologies such as Python, machine learning libraries (TensorFlow, Keras, PyTorch, Scikit-learn), image processing libraries (OpenCV, PIL), and potentially encryption libraries (PyCryptodome). The key features of the system will be automated skin disease prediction, image-based analysis, and a user-friendly interface, with the overarching benefit of improving the accuracy and efficiency of skin disease diagnosis, enabling potential for early detection, supporting clinicians, and increasing access to dermatological expertise.

III. LITERATURE SURVEY :

The literature survey in the provided document focuses on existing approaches to image encryption. It highlights various techniques and algorithms used to secure image data.

- One study presents a new chaotic key-based design for image encryption and decryption, proposing a VLSI architecture for this algorithm. This approach uses bitwise XORed or XNORed operations to predetermine keys for the chaotic binary sequence of the gray level of each pixel. It emphasizes features like low computational complexity, no distortion, and high security. The VLSI architecture is noted for its advantages, such as low hardware cost, high computing speed, and hardware utilization efficiency.
- Another study presents a modified AES-based algorithm for image encryption. It acknowledges that encryption is a common technique for providing security for images, which have wide applications in internet communication, multimedia systems, medical imaging, telemedicine, and military communication. The survey also mentions different image protection techniques, such as vector quantization, where the image is decomposed into vectors for encoding and decoding, or by dividing the image into a large number of shadows to make it undetectable to illegal users.
- Some authors present secure image encryption using AES, highlighting security as a major issue. With the increased transmission of images for communication, providing confidentiality from unauthorized access is a primary concern. AES is used for encryption and decryption, converting the image into an unrecognizable form using a key, and later converting it back to the original image by an authorized receiver.
- Further research focuses on image encryption and decryption using the AES algorithm, emphasizing its role in effectively securing image communication. AES has replaced the Data Encryption Standard (DES) by offering enhanced security. AES key expansion uses the 128-bit key for the encryption process by using bitwise exclusive OR operation of image set pixels.
- One study presents an image encryption based on AES key expansion, noting the specific characteristics of images, such as high transmission rates with limited bandwidth, redundancy, bulk capacity, and correlation among pixels, which must be considered during encryption. The AES algorithm with key expansion is used, where the encryption process involves a bitwise exclusive OR operation of image pixels set along with a 128-bit key. The key is generated at the sender and receiver side based on the AES Key Expansion.

Aims And Objectives :

The primary aims of a skin disease prediction project using machine learning typically revolve around these objectives:

- **Develop an automated system for accurate and efficient skin disease prediction:** This involves creating a system that can analyze images of skin lesions and predict the likelihood of various skin diseases, reducing reliance on manual and potentially subjective assessments.
- **Assist clinicians in diagnosis and improve patient care:** The system aims to serve as a supportive tool for healthcare professionals, aiding in decision-making, increasing diagnostic accuracy, and enabling earlier detection and intervention for skin conditions.
- **Enable early detection and screening of skin diseases:** By automating the analysis of skin images, the system can facilitate large-scale screening and early detection of potentially serious skin conditions, leading to better patient outcomes.
- **Enhance accessibility to dermatological expertise:** In areas with limited access to specialists, an automated system can provide valuable support for primary care physicians and other healthcare providers in identifying potential skin problems.
- **(If applicable) Ensure secure handling of patient data:** If the system processes and stores sensitive patient information, an aim may be to implement robust security measures to protect patient privacy and confidentiality.

Objectives:

- Data: Acquire, prepare, and preprocess image datasets.
- Features: Extract and select relevant image features.
- Model: Develop, train, evaluate, and validate ML models.
- System: Implement a user-friendly prediction system.
- Security: (If needed) Ensure data security and privacy.

Methodology :

1. Data Acquisition and Collection:

- **Dataset Identification:** Identify suitable publicly available datasets of skin disease images (e.g., HAM10000, ISIC Archive) or establish collaborations with medical institutions to acquire data.
- **Data Gathering:** Collect images representing a diverse range of skin diseases, ensuring sufficient examples for each category.
- **Data Annotation:** Verify the accuracy of existing annotations or, if necessary, annotate the images with accurate diagnoses by qualified dermatologists.
- **Data Splitting:** Divide the dataset into three subsets:
    - **Training Set:** Used to train the machine learning models.
    - **Validation Set:** Used to tune hyperparameters and evaluate model performance during training.
    - **Test Set:** Used to assess the final performance of the trained models.

2. Data Preprocessing:

- **Image Resizing and Normalization:** Resize images to a consistent size and normalize pixel values to a specific range (e.g., 0-1) to ensure uniformity and improve model convergence.
- **Noise Reduction:** Apply filtering techniques (e.g., median filtering) to reduce noise and artifacts in the images.
- **Image Enhancement:** Enhance image quality using techniques like contrast stretching, histogram equalization, or color correction to improve feature visibility.
- **Data Augmentation:** Generate additional training data by applying transformations to existing images, such as rotation, flipping, zooming, or cropping, to increase dataset size and diversity and improve model robustness.

3. Feature Extraction:

- **Feature Extraction Techniques:** Employ appropriate techniques to extract relevant features from the preprocessed images:
    - **Manual Feature Engineering:** Extract features based on domain expertise, such as color histograms, texture features (e.g., Haralick features), shape features, or lesion size and border characteristics.
    - **Automated Feature Extraction:** Utilize deep learning models, such as Convolutional Neural Networks (CNNs), to automatically learn hierarchical representations of image features.
- **Feature Selection/Dimensionality Reduction:**
    - Apply feature selection methods (e.g., chi-square test, mutual information) or dimensionality reduction techniques (e.g., Principal Component Analysis (PCA)) to select the most informative features and reduce the complexity of the model.

4. Machine Learning Model Development:

- **Model Selection:** Choose appropriate machine learning algorithms for skin disease prediction:
    - **Convolutional Neural Networks (CNNs):** If using deep learning for feature extraction and classification.
    - **Support Vector Machines (SVMs):** If using manual feature engineering.
    - **Random Forest:** An ensemble method suitable for various feature types.
    - Other algorithms like k-Nearest Neighbors (KNN) or Naive Bayes can also be explored.
- **Model Architecture Design:**
    - If using CNNs, design the architecture of the network, including the number of layers, filter sizes, activation functions, and pooling layers.
- **Model Training:**
    - Train the selected machine learning models using the training dataset.
    - Optimize model parameters using appropriate optimization algorithms (e.g., Adam, SGD).
    - Monitor model performance on the validation set during training to prevent overfitting and tune hyperparameters.
- **Hyperparameter Tuning:**
    - Adjust hyperparameters (e.g., learning rate, batch size, number of epochs, regularization parameters) using techniques like grid search or cross-validation to optimize model performance.

5. Model Evaluation and Validation:

- **Performance Metrics:** Evaluate the trained models on the test dataset using appropriate metrics:
    - **Accuracy:** Overall correctness of predictions.
    - **Precision:** Proportion of correctly predicted positive cases.

- **Recall (Sensitivity):** Proportion of actual positive cases correctly predicted.
- **Specificity:** Proportion of actual negative cases correctly predicted.
- **F1-score:** Harmonic mean of precision and recall.
- **Area Under the ROC Curve (AUC):** Measure of the model's ability to distinguish between classes.

- **Cross-Validation:**
  - Perform k-fold cross-validation to assess the model's generalization ability and robustness.
- **Confusion Matrix Analysis:**
  - Analyze the confusion matrix to identify specific areas where the model performs well or struggles.

6. System Implementation and Deployment:

- **Software Development:** Develop a user-friendly interface or application that allows users (e.g., clinicians) to input skin images and receive predictions.
- **Integration:** Integrate the trained model into the application or system.
- **Deployment:** Deploy the system in a clinical setting or make it accessible through a web platform or mobile app.

7. Security and Privacy (If Applicable):

- **Data Encryption:** Implement encryption techniques to protect sensitive patient data during storage and transmission.
- **Access Control:** Establish access control mechanisms to ensure that only authorized personnel can access patient data and the prediction system.
- **Data Privacy Compliance:** Ensure compliance with relevant data privacy regulations (e.g., HIPAA, GDPR)

REFERENCES:

1. M. Zeghid, M. Machhout, L. Khriji, A. Baganne, and R. Tourki , "A Modified AES Based Algorithm for Image Encryption" ,2007 .
2. P. Radhadevi, P. Kalpana , "Secure Image Encryption Using Aes",2012 .
3. Roshni Padate, Aamna Patel, "Image Encryption And Decryption Using Aes Algorithm" ,2014.
4. Jose´ J. Amador, Robert W. Green, "Symmetric-Key Block Cipher for Image and Text Cryptography", 2005.
5. Philip P. Dang and Paul M. Chau, "Image Encryption For Secure Internet Multimedia Applications", 2000.