Optimized Retrieval and Secured Cloud Storage for Medical Surgery Videos Using Deep Learning

Megala G[1], Swarnalatha P[2]*
Research Scholar[1], Professor[2]
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India 632014[1,2]

***Abstract*—Efficient secured storage and retrieval of medical surgical videos are essential for modern healthcare systems. Traditional methods often struggle with scalability, accessibility, and data security, necessitating innovative solutions. This study introduces a novel deep learning-based framework that leverages a hybrid algorithm combining a Variational Autoencoder (VAE) and Group Lasso for optimized video feature selection. This approach reduces dimensionality and enhances the retrieval accuracy of video frames. For storage and retrieval, the system employs a weighted graph-based prefetching algorithm to manage encrypted video data on the cloud, ensuring both speed and security. To ensure data security, video frames are encrypted before cloud storage. Experimental results show that this system outperforms current methods in retrieval speed and accuracy of 99% while maintaining data security. This framework is a significant advancement in medical data management, offering potential applications across other fields that require secure handling of large data volumes.**

***Keywords*—Medical video storage; feature selection; Variational Auto Encoder (VAE); weighted-graph-based prefetching algorithm; group lasso**

I. Introduction

The increasing demand for effective management of medical surgical videos is a pressing challenge in the healthcare sector. These videos are crucial for various purposes, including surgical training, research, and patient care. Traditional methods of storing and retrieving medical video data face significant limitations related to scalability, accessibility, and security. As healthcare facilities generate vast volumes of video data daily, there is a growing need for solutions that can efficiently handle these data while ensuring data integrity and confidentiality.

The process of storing and retrieving recordings of medical procedures, such as surgery, is an essential duty in the field of healthcare [1]. This is because these movies include information that is beneficial to both medical personnel and researchers [2]. Unfortunately, conventional approaches to the storage and retrieval of medical video data have limitations, both in terms of their capacity to manage vast volumes of data and of their ability to guarantee the confidentiality of the data [3], [4], [5].

Recent advancements in deep learning have shown great promise in addressing some of these challenges, particularly in the field of medical image analysis. However, the application of deep learning techniques to medical video data is still in its nascent stages, presenting a significant research gap. Current systems often fail to optimize both the retrieval speed and the security of stored data, leading to inefficiencies and potential vulnerabilities in data management.

Medical image analysis is an area where deep learning techniques have shown significant promise in recent years [8]. These deep learning techniques are used to increase the efficacy of medical video storage and retrieval, as well as the security of these processes [6], [7]. Using a system that is based on deep learning, this research presents a unique method for the safe storage of medical operation footage in the cloud, with the goal of optimizing the retrieval of certain moments [8], [9], [10].

This paper addresses this gap by proposing a novel deep learning-based framework designed to enhance the storage and retrieval of medical surgical videos on cloud platforms. By integrating a hybrid algorithm combining a Variational Autoencoder (VAE) and Group Lasso, the proposed approach aims to optimize feature selection and improve the accuracy of video retrieval. Furthermore, the use of a weighted graph-based prefetching algorithm for encrypted video storage enhances both the security and speed of data access.

The proposed system is comprised of three primary modules: one that converts video into frames, another that divides frames into sets, and a third that employs a hybrid approach that makes use of a variational autoencoder, group lasso, and feature selection through the application of a weighted-graph-based prefetching algorithm [11]. The usage of a variational autoencoder helps in lowering the dimensionality of the video frames, and the group lasso approach helps in selecting the features that provide the most useful information about the scene [12]. The weighted-graph-based prefetching technique is responsible for the improved retrieval performance [13]. It does this by anticipating which frames are going to be retrieved the most often [14]. These methods are included into a framework that is based on deep learning, which enables the system to learn and improve the feature representations that are derived from the video frames [15].

Priyanka et. al. [16] reviewed encryption methods for securing medical data, emphasizing the importance of confidentiality and integrity in healthcare applications. Encrypting the frames prior to saving them in the cloud [17], [18] is one of the ways that the suggested method protects the confidentiality of the patient information. Using methods that are based on deep learning makes it possible to store and retrieve vast volumes of medical video data in an effective manner while still ensuring the confidentiality of the data [19]. The suggested system is used to enhance the efficiency and security of the storage and retrieval of medical footage, and it also has the potential to find applications in other fields where huge volumes of data need to be kept and retrieved in a safe manner [20]. Overall, the proposed system offers a promising solution to the challenges of medical surgery video storage and retrieval [21] and demonstrates the potential of deep learning-based approaches in addressing complex healthcare problems.

A. Problem Formulation

Despite the advancements in medical data management, existing systems for storing and retrieving surgical videos face several critical challenges such as scalability, data security, feature extraction and retrieval accuracy. Traditional storage solutions are often inefficient at handling the ever-increasing volume of medical video data, leading to slower retrieval times and higher storage costs. Ensuring the security of sensitive medical data is paramount, yet current systems often lack robust mechanisms for protecting data during storage and transmission. Existing techniques may not fully utilize the potential of video data, leading to suboptimal feature extraction and retrieval accuracy.

B. Major Contributions

The major contributions of this research work are:

- A novel hybrid algorithm that combines a Variational Autoencoder (VAE) with Group Lasso is proposed to effectively reduce video frame dimensionality and select the most informative features. This approach significantly improves retrieval accuracy.
- The proposed approach employs a weighted graph-based prefetching algorithm for encrypted video storage and retrieval. This method optimizes data retrieval speed while ensuring robust security measures for protecting sensitive medical information.
- The developed application is deployed on AWS cloud infrastructure to ensure scalability and reliability.

By addressing these key challenges, the proposed framework offers a comprehensive solution for managing medical surgical videos and has the potential to be applied across various domains requiring secure and efficient data handling.

The structure of this article is as follows. Several authors address several strategies for safely archiving medical surgery footage that are discussed in Section II. Section III displays the proposed framework. Section IV details the investigation's results. In Section V discusses the conclusion with limitation and future work.

II. Related Works

Al Abbas et al. [1] discuss the benefits of a surgical video collection in the academic surgical context, as it can aid in preoperative preparation, mentoring of medical students, and research into surgical technique and expertise. Chen, Y. et al. [3] highlight the challenges of data dispersion in the medical field and propose using blockchain as a secure and accountable supply chain for storing and exchanging medical data. Khelifi, F. et al. [22] present a solution for securely storing and exchanging data in the cloud that does not rely on RDH, which has been deemed ineffective in such applications. Li, Y. et al. [7] propose the Security-Aware Efficient Distributed Storage (SA-EDS) paradigm to prevent cloud service providers from accessing private customer data stored in the cloud. Nguyen, D. et al. [23] introduce a novel EHR sharing strategy that combines mobile cloud computing and blockchain to securely and efficiently share medical data across mobile cloud settings.

Prachi Deshpande et al. [24] introduce WCDM, a watermark compression and decompression module that can ensure the safety of video files stored in the cloud by adding a watermark based on block-by-block calculation and then using data compression. Srivastava, P., & Garg, N. [13] discuss the potential of IoT and the need for collaboration across research groups to address the challenges posed by IoT-related issues [25]. Usman, M. et al. [15] propose a secure approach to encrypting hidden data in compressed video streams, which is essential to address concerns about privacy and security in public clouds. Yang, Y. et al. [20] suggested a health IoT data management where security is ensured by implementing a lightweight distributed access control system that makes advantage of quick keyword searching. It reduces the computational burden on low-powered IoT devices by allowing for remote trapdoor generation, encryption, and decryption.

Convolutional Neural Networks (CNNs) have significantly impacted medical imaging and video analysis due to their ability to automatically learn complex features from data. They have been applied to tasks such as image classification, segmentation, and anomaly detection. For instance, Kolarik et al. [26] conducted a comprehensive survey on deep learning in medical imaging, emphasizing the versatility of CNNs in enhancing the analysis of medical images and videos. Their work highlights how CNN architectures, such as AlexNet and VGG, have been adapted for medical video data to identify surgical tools and assess procedure quality.

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been utilized for capturing temporal dependencies in video sequences. Sánchez-Caballero et al. [27] introduced Long-term Recurrent Convolutional Networks (LRCNs) for activity recognition in videos, demonstrating how integrating CNNs with RNNs can effectively handle the temporal dynamics present in surgical videos. This integration is crucial for tasks such as predicting surgical phases and detecting anomalies during procedures.

Secure Cloud Storage Frameworks leverage encryption and advanced access control mechanisms to safeguard medical data in cloud environments. Biksham et al. [28] proposed a framework that employs homomorphic encryption, allowing computations on encrypted data without revealing the actual content. This approach ensures data privacy while enabling efficient processing and retrieval, crucial for healthcare applications that require both security and functionality. Cloud-based Healthcare Systems offer scalable and cost-effective solutions for managing medical data. Islam et al. [29] explored the integration of cloud computing in healthcare, emphasizing benefits such as remote accessibility, resource scalability, and data backup. They discussed how cloud services can facilitate real-time collaboration among healthcare professionals, enhancing patient care and research capabilities.

AWS Cloud Services for Healthcare provide a robust infrastructure for deploying healthcare applications. Amazon offers various services offered by AWS, including secure storage solutions, machine learning platforms, and compliance with healthcare regulations such as HIPAA. These services enable healthcare providers to deploy scalable and secure applications, supporting the growing demand for efficient data management. Hybrid Approaches for Video Retrieval combine content-based and text-based methods to enhance retrieval accuracy. Unar et al. [30] proposed a hybrid approach that leverages both visual and textual features for video retrieval, demonstrating improved performance in retrieving relevant video content. Such approaches are beneficial for medical video retrieval, where both visual and contextual information play a critical role.

The discussed studies provide a strong foundation for developing optimized retrieval and secure storage solutions for medical surgical videos. The proposed framework in this research leverages these advancements, aiming to address scalability, security, and efficiency challenges in managing medical video data.

III. Materials and Methods

The proposed method involves a comprehensive framework for securely storing medical surgical videos in the cloud and efficiently retrieving them using a given image query. The framework is designed to address the challenges of data security, retrieval speed, and accuracy. It consists of three main components: feature extraction and selection, secure storage, and image query-based retrieval. The proposed approach uses a deep learning-based framework, which includes a variational autoencoder with group lasso for hybrid feature selection and a weighted-graph-based prefetching algorithm to improve security and retrieval speed. Before being uploaded to the cloud, the frames are encrypted to protect the confidentiality of the patient's information. Results from experiments show that the suggested system improves upon state-of-the-art methods without compromising the privacy of patient's medical records during retrieval.

A. Frame Conversion

Each surgical video is divided into frames, and key frames are identified based on changes in visual content and motion. This reduces the amount of data while preserving important information. The process of converting a video into frames involves extracting each individual frame from the video and saving it as a separate image file. This process is represented mathematically using the following equation

$$F(i, j, k) = V(i, j, k) \tag{1}$$

where F is the resulting frame image, V is the original video, and i, j, and k are the frame, row, and column indices, respectively. This equation represents the process of extracting the ith frame from the video V at row j and column k and storing it as a separate image file. Once the video has been converted into frames, the frames are divided into sets for efficient storage and retrieval. This is done by grouping the frames together based on their similarity or temporal proximity. The exact method used for dividing the frames into sets can vary depending on the specific application and requirements.

B. Feature Selection using Hybrid Method

After the Frame Conversion, a feature selection technique is applied to identify relevant and informative features from the extracted frames. A combination of Variational Autoencoder (VAE) and Group Lasso is used to extract and select informative features from the video frames. The VAE reduces dimensionality by learning compact feature representations, while Group Lasso selects the most relevant features, minimizing noise and redundancy.

A variational autoencoder (VAE) is a type of neural network that can learn a low-dimensional representation of high-dimensional data such as images or videos. In the proposed methodology, the VAE is used to reduce the dimensionality of the video frames, which makes them easier to store and retrieve. The VAE works by mapping the high-dimensional video frames to a lower-dimensional space, where each dimension corresponds to a specific feature. Group lasso is a feature selection technique that works by selecting groups of features rather than individual features. In the proposed methodology, group lasso is used to select the most informative groups of features from the lower-dimensional space generated by the VAE. This is important because not all features are equally important for the optimal storage and retrieval of medical surgical videos.

The VAE is inherently nonlinear because it involves encoding the input through layers of neural networks that apply non-linear ReLU activation functions to map the input into a lower-dimensional latent space. Group Lasso technique is used for regularization to select groups of features while reducing dimensionality. The operation of group lasso involves optimization, which is also non-linear. Hence combining both involves non-linear operation. It captures the complex, non-linear relationships in the video data across the spatial and temporal dimensions. The matrix $V$is subjected to the VAE, which will encode the frame into a latent representation, and group lasso will regularize this representation by selecting the most informative features.

1) Variational auto encoder: After applying a feature selection technique Variational Autoencoder is employed, which is a type of neural network model, for unsupervised feature learning. A Variational Autoencoder is a generative model that consists of two main components: an encoder and a decoder as shown in Fig. 1. The encoder takes in input data, in this case, the extracted frames from the video, and maps them to a latent space representation. The latent space is a lower-dimensional representation that captures the underlying structure and patterns in the input data. During the training phase, the VAE aims to learn the probability distribution of the latent space that best explains the input data. It does so by maximizing the evidence lower bound (ELBO), which is a measure of how well the model reconstructs the input data while simultaneously regularizing the latent space.

Data preprocessing involves preparing the data for the model. The data needs to be normalized, and any missing values need to be handled appropriately. The encoder network takes the input data and produces a set of latent variables. This network is typically composed of one or more fully connected layers with activation functions such as ReLU or Sigmoid. The GAN network takes the output of the encoder network and generates a set of images. Both a generator and a discriminator network make up the GAN system. Images are created by the generator network using the latent variables, and the discriminator network compares the produced pictures to

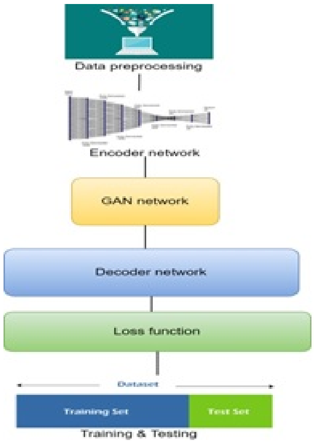


Fig. 1. Variational autoencoder.

the real world counterparts. The decoder network takes the output of the GAN network and produces a reconstruction of the input data. The decoder network is composed of two fully connected layers with ReLU activation functions. The loss function is a hybrid of the reconstruction loss and the GAN loss which is employed to compare the original data to the rebuilt version.

Backpropagation and stochastic gradient descent are used to train the model. To improve the quality of the reconstructed data and the generated images, the model is trained to minimize the loss function. Once the model is trained, it is used to generate new data by sampling the latent variables from a distribution and passing them through the decoder network. The quality of the generated data is evaluated by comparing it to the original data and measuring the reconstruction loss.

Assume N frames or images $X^{(n)}{}_{n=1}$, with $X^{(n)} \in R^{N_x \times N_y \times N_c}$; Where N_x and N_y represents the spatial dimension and N_c represents the number of color channels. The simplest possible configuration of the decoder with 2 layers (L) is used in generative model. the data generation is shown in Eq. 2; the definitions of Layer 1, unpooling and layer 2 is shown in Eq. 3, Eq. 4 and Eq. 5, respectively.

$$X^{(n)} \approx n(S^{(n,1)}, \alpha_0^{-1} I) \tag{2}$$

$$S^{(n,1)} = \sum_{k_1=1}^{k_1} D^{(K_1,1)} * S^{(n,k_1,1)} \tag{3}$$

$$S^{(n,1)} \approx unpool(S^{(n,2)}) \tag{4}$$

$$S^{(n,2)} = \sum_{k_2=1}^{k_2} D^{(K_2,1)} * S^{(n,k_2,1)} \tag{5}$$

To clarify the notation, 3D tensors are denoted by expressions with two superscripts, such as $D^{kl,l}$ and $S^{(n,l)}$ for layers $l \in \{1, 2\}$. The $S^{(n,kl,l)}{}_{kl=1}^{kl}$ is stacked in space to produce the tensor $S^{(n,l)}$. Each of the Kl1 2D slices of the 3D $D^{kl,l}$ is convolved with the 2D spatially-dependent $S^{(n,kl,l)}$ in the convolution $D^{kl,l}S^{(n,kl,l)}$; by aligning and stacking these convolutions, a tensor output is seen for $D^{kl,l}S^{(n,kl,l)}$.

The VAE is a latent variable model in which x represents the set of observable variables, z represents the set of stochastic latent variables, and $p_x \times p_y$ represents a parameterized model of the joint distribution. The goal is to maximize the average marginal log-likelihood for the given dataset. Yet, when neural networks are used to parameterize the model, it is often impossible to suppress this expression. The problem is solved, in part, by using variational inference to optimize the evidence lower Limit (ELBO) for each observation is shown in Eq. 6.

$$\log p(x) = \log p(x,z) d_z \geq E_{q(z)}[\log p(x|z)] - KL(q(z)||p(z)) \tag{6}$$

The variational family contains the approximate posterior $q(z)$. The ultimate goal is achieved by introducing an inference network $q(z|x)$ that gives a probability distribution for each data point x.

$$l(x;\theta) = E_{q(z|x)}[\log p(x|z)] - KL(q(z|x)||p(z)) \tag{7}$$

Using the re-parameterization trick $q(z|x)$, Monte Carlo estimation is used to efficiently approximate the ELBO for continuous latent variable z.

The dataset x with n independent and identically distributed samples are generated by a latent variable z that represents the ground truth. Let $p(x|z)$ stand for a neural network's probabilistic decoder, which generates x from z in the presence of uncertainty. The neural network-based encoder produces a variation posterior, $q(z|x)$, which is a close approximation to the distribution of representation for dataset x. In the realm of creating models, the Variational Autoencoder (VAE) is a popular choice. The core idea behind VAE is stated as follows: (1) VAE employs a probabilistic encoder whose parameters are determined by a neural network to produce a latent variable z as a distributional representation of the input data samples x. (2) Next, run z samples through the decoder to get back the original input data. Maximizing the marginal probability of the reconstructed data is the goal of VAE, however the method also involves intractable posterior inference. To maximize its variational lower limit log likelihood, researchers apply methods like backpropagation and stochastic gradient descent.

$$\log P_\theta(x) \geq L_u ae = E_{q_\emptyset(z|x)}[\log P_\theta(x|z)] - D_K L(q_\emptyset(z|x)||p(z)) \tag{8}$$

where z is the random variable and $p(z)$ is the prior distribution (often a Gaussian). Both the variational posterior, $q(z|x)$, and the probabilistic decoder, $p(x|z)$, are parameterized by a neural network to create data x given the latent variable z. On training the data in hybrid VAE, KL terms are used for reconstruction of terms to avoid vanishing gradients.

The overall architecture of securely storing videos in cloud and retrieving videos is shown in Fig. 2. The input video for this process is being converted into a set of frames. This process involves breaking down the video into individual frames, essentially creating a series of still images that are used for a variety of purposes such as video analysis, motion tracking, and image processing. The conversion process typically involves opencv libraries to extract frames from the video file. These frames are then saved as individual image files, with each frame representing a moment in time from the original video. Once the frames are extracted, they are processed in

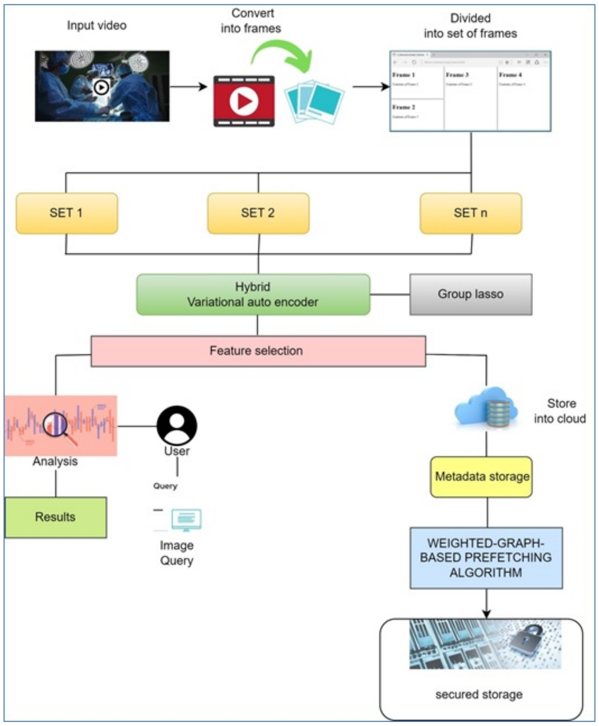


Fig. 2. Overall framework for secured video storage and retrieval.

various ways, such as filtering, resizing, or adjusting color balance. This is useful for a wide range of applications, from analyzing the movement of objects in the video to create visual effects.

2) Group-Lasso: After utilizing a Variational Autoencoder the Group-Lasso technique is then used to further refine the chosen characteristics. The Group-Lasso method as shown in Fig. 3 is a regularization approach that encourages feature sparsity. It is especially effective when working with high-dimensional data, such as video frames, where the number of features (pixels) are enormous. The Group-Lasso method is used to learn the latent space representation derived from the VAE in the context of video analysis. This latent space representation captures the video frames' most informative and significant properties.

The first step is to preprocess the data. This includes standardizing the data, normalizing it, inputting missing values, and feature scaling. The information is divided into a training set and a test set. The model is fit to the training set (70%), and its efficacy is tested on a separate test set (30%).The features are grouped based on their domain knowledge. Features that have similar characteristics or related to each other are grouped together. A Group-Lasso model is fitted to the training data. The objective function is the sum of the squared errors plus the regularization term. The regularization term is a combination of the L1 and L2 norms of the regression coefficients. The regularization parameter are tuned to obtain the best model performance. This is done by using cross-validation or grid search.Once the model is trained, the features that have non-zero coefficients are selected as important features. This allows for feature selection and grouping to be performed simultaneously. The final step is to evaluate the model performance on the test data. This is done by computing metrics such as the mean squared error, R-squared, and the coefficient of determination. The performance is compared to other models to determine which one is the best fit for the data.

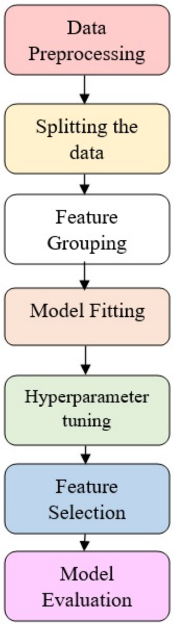


Fig. 3. Group Lasso.

This part builds upon off, but with the addition of material on the Group-Lasso problem and a bigger class of probability functions. A generalized linear model (GLM) has three components, as stated in Eq. 9.

$$f(y,\theta,\emptyset) = Exp(\emptyset^{-1}(y\theta - b(\theta)) + c(y,\emptyset)) \tag{9}$$

In Eq. 9, the average response, $E_\theta[y]$, is related to the free parameter through the formula $\mu = b(\theta)$. The link function g is strictly monotone differentiable function. However, it is restricted to canonical link functions when $g(\mu) = \tau = \theta$. Therefore, the parameterization $f(y;\theta;\emptyset)$ is done.This framework is technically distinguished by the strict concavity in of the function $\log f(y;\theta;\emptyset)$. In order to have minimal one-dimensional acceptable statistics, y, the log partition function $b(\theta)$ must be strictly convex.

The normal distribution with mean $\tau = \mu$ and variance $b(\tau) = (1/2)\tau^2$ is a specific instance from which the conventional linear regression model is formed. The goal of this section is to introduce the issue of minimizing the negative log-likelihood given an independent, identically replicated data sample $x_1, ..., x_n$, organized as rows of the data matrix X, and a corresponding vector of replies $y = (y_1, ..., y_n)^T$.

$$l(y,n,\emptyset) = -\sum_i \log f(y_i; n_i; \theta) = \sum_i \emptyset^{-1}(y_i n_i - b(n_i)) + c(y_i,\theta) \tag{10}$$

$$\nabla_{n^l}(n) = -(y - g^{-1}(n)) \tag{11}$$

$$\nabla_{\beta^l}(\beta) = -X^T \nabla_{n^l}(n) = -X^T(y - g^{-1}(X\beta)) \tag{12}$$

Hessians are computed using Eq. 13

$$H_n = W, H_\beta = X^T W X \tag{13}$$

The linear model estimation is shownn in Eq. 10. It involves minimizing the mean square error risk, which is defined as in Eq. 14.

$$L_{ls}(\beta) = ||Y - x\beta||_2^2 \tag{14}$$

$L_{ls}(\beta)$ is a np matrix whose rows include the feature vectors $x_i i = 1, \ldots 1_n$ and $Y = [Y, ..., Yn]$. In the derivation and characteristics of the least squares estimate, the matrix(β), also known as the design matrix, plays a crucial role. In reality, the minimize of $L_{ls}(\beta)$ is known if $n > p$. Since the matrix is ill-conditioned and lacks the correct inverse, the estimate LS is not unique and does not even exist when dealing with high-dimensional data. A simple replacement of $||x\beta||_2^2$ with $||Y - x\beta||_2^2$, where $\beta \geq 0$ is the regularization parameter, would regularize the matrix. The ridge regression loss function provides a formalization of this method, as shown in Eq. 15.

$$L_{Ridge}(\beta; y) = L_{ls}(\beta) + y||\beta||_2^2 \tag{15}$$

Parameterized by Eq. 15, the ridge regression is defined as the minimizer of L_{Ridge}. cross-validation variant is used to assess predictive risk allows for the optimal selection. Regularization of the conventional least squares solution (ls) is provided by the ridge regression solution, although the reduced complexity solution cannot be generated in the situation of high dimensional feature vectors. The estimate of the j^{th} component is obtained from the ridge regression solution which is in fact a non-negative.

$$L_{lasso}(\beta; y) = L_{ls}(\beta) + y||\beta|| \tag{16}$$

As the risk function is minimized, the Lasso regression estimator is established. A few components of it is made zero using the $L1$-penalty based method. A value of 0 for the estimated j indicates that the J^{th} feature contributes nothing to the model's prediction ability and is left out. As a result, Lasso enables both the estimation and model selection at the same time.

3) Hybrid feature selection method: The hybrid feature selection method combines the use of a variational autoencoder (VAE) and group lasso to select informative features from the input data. The VAE is trained on the video frames to learn a low-dimensional representation of the frames, which is expressed as shown in Eq. 17.

$$z = Encoder(x; \theta) z = Encoder(x; \theta) \tag{17}$$

Next, group lasso is applied to the low-dimensional representation to select the most informative groups of features. This is achieved by solving the following optimization problem as shown in Eq. 18.

$$\begin{gathered}\beta = \arg\min \beta 12n||y - X\beta||22 + \lambda \sum j = 1pwj||\beta j||2 \\ \beta = \arg\min \beta 2n1||y - X\beta||22 + \lambda \sum j = 1pwj||\beta j||2\end{gathered} \tag{18}$$

where $\beta 12n$ is the output variable, $||y - X\beta||$ is the input variable, $\arg\min \beta 2n1$ is the coefficient vector, $|n|$ is the sample size, (p) is the number of features, (pwj) is a weight assigned to the $|j|^{th}$ feature, and a tuning parameter that controls the strength of the penalty term. The solution vector $\beta = [\beta_1, \beta_2, \ldots \beta_p]$ contains the selected features. The resulting feature representation is then used for storage and retrieval of the video frames, allowing for improved efficiency and accuracy in handling large amounts of medical video data. The hybrid feature selection method works by training the VAE on the video frames to learn a low-dimensional representation of the frames. Then, group lasso is applied to the low-dimensional space to select the most informative groups of features. The resulting feature representation is then used for storage and retrieval of the video frames. Overall, the hybrid feature selection method is an effective way to select the most informative features from medical surgical videos for optimal storage and retrieval. The use of a VAE and group lasso allows for efficient dimensionality reduction and feature selection, respectively, resulting in a more efficient and accurate system.

C. Video Data Security Using a Weighted-Graph-Based Prefetching

Following the feature selection method, a data security method is implemented based on a Weighted-Graph-Based Prefetching technique. This method seeks to safeguard video data by automatically prefetching and storing encrypted video chunks depending on their relevance and access patterns.

The proposed system employs a Weighted-Graph-Based Prefetching algorithm for video data security. This algorithm predicts the most likely frames to be accessed based on previous access patterns and prefetches them, reducing the time needed to retrieve data and improving system performance. The algorithm works by constructing a graph of video frames, where each frame is represented as a node in the graph. The edges between nodes are weighted based on the similarity between the frames. This similarity is calculated using a feature vector that is extracted from the frames using the hybrid feature selection method described earlier.

Video data security is a critical aspect of any system that deals with sensitive medical data. To ensure that the medical video data is secure, the proposed system uses encryption to protect the frames before they are stored in the cloud. Specifically, each video frame is encrypted using an encryption key to prevent unauthorized access to the data. The encryption process is as follows

$$EncryptedFrame = Encrypt(Frame, Key) \tag{19}$$

where $Encrypt$ is a AES cryptographic function that takes the video frame and encryption key as inputs and produces an encrypted frame as output. Before storing in the cloud, video frames are encrypted using advanced encryption algorithms like AES (Advanced Encryption Standard) to ensure data confidentiality and integrity. Encrypted video frames are stored in a cloud infrastructure, such as AWS, using a weighted graph-based prefetching algorithm. This algorithm predicts the most likely frames to be accessed, optimizing storage retrieval and enhancing data security.

In addition to encryption, the proposed system uses a Weighted-Graph-Based Prefetching algorithm to optimize the retrieval of the video frames from the cloud while maintaining security. The algorithm works by predicting the most likely frames to be accessed and retrieving and caching them in advance for improved performance and responsiveness. The algorithm takes into account the frequency of access and the distance between frames to determine the weights of the edges in the graph.

The Weighted-Graph-Based Prefetching algorithm is based on the following Eq. 20:

$$P(i) = \alpha * R(i) + (1-\alpha) * \sum(j \in N(i)) w(i,j) * R(j) \quad (20)$$

where $P(i)$ is the predicted probability of frame i being accessed, $R(i)$ is the frequency of access of frame i, $N(i)$ is the set of neighboring frames of i,$w(i,j)$ is the weight of the edge connecting frames$i$ and j, and α is a damping factor that balances the contribution of $R(i)$ and $\sum(j \in N(i))w(i,j) * R(j)$ to the prediction.

The weights of the edges between frames are determined based on the distance between frames and the frequency of access. The distance between frames is calculated using the following Eq. 21.

$$d(i,j) = ||f(i) - f(j)||^2 \quad (21)$$

Where f (i) and f (j) are the feature representations of frames i and j, respectively, which are obtained through the hybrid feature selection method described earlier.

The weight of the edge between frames i and j is then calculated using the following Eq. 22:

$$w(i,j) = exp(-d(i,j)/\sigma) \quad (22)$$

where σ is a scaling factor that controls the influence of the distance on the weight. By combining encryption and the Weighted-Graph-Based Prefetching algorithm, the proposed system ensures the security and privacy of medical video data while optimizing the retrieval process for improved performance and responsiveness. Overall, the Weighted-Graph-Based Prefetching algorithm is an effective way to improve the security and retrieval speed of medical surgical videos. By predicting and prefetching the most likely frames to be accessed, the algorithm can reduce retrieval time and improve system performance. Additionally, by encrypting the data and monitoring access, the system ensures the security and confidentiality of the medical data.

D. Image Query-Based Retrieval

When an image query is received, its features are extracted using the same hybrid feature selection method. These features are then matched against the stored video frame features using similarity measures like cosine similarity. The weighted graph-based prefetching algorithm helps efficiently retrieve video frames by leveraging pre-computed weights that represent the likelihood of frame access based on historical access patterns and feature similarity. The proposed secured video storage and retrieval framework is illustrated in the Algorithm 1.

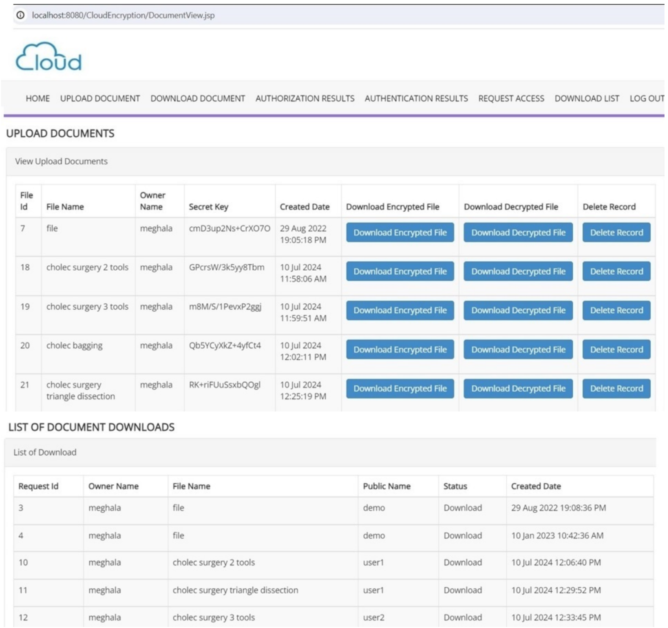
Algorithm 1 Proposed algorithm

```
Input: Cholec80 surgical videos $V$, Image query $Q$
Output: Encrypted video stored in the cloud and Retrieved video relevant to $Q$
Divide each video $V$ into a set of video frames $F = f_1, f_2, ...f_n$
Hybrid feature selection
for $F_i = 1$ to $n$ do
    Apply VAE to extract features by learning a low-dimensional representation $Z_i$ of the video frames in $F_i$
    Apply Group Lasso to each $Z_i$ to select features set $S_i$.
    Concatenate the selected feature sets to obtain the final feature set $S$.
end for
Encrypt selected Feature set $(S')$ and store metadata in cloud.
Build a weighted graph $G$ whose nodes correspond to the subsets $F_i$ and whose edges correspond to the probability that a subset is accessed after another.
for all Video Request do
    Decrypt $S$
    Compute similarity score between $Q$ and $S$.
    Prioritize frame retrieval with high similarity scores using weighted-graph-based prefetching algorithm
end for
return Output the retrieved video segments relevant to $Q$.
```

The use of VAE and Group Lasso ensures that only the most informative features are retained, enhancing retrieval accuracy by focusing on the most relevant aspects of the video content. Encrypting video frames before storage guarantees data confidentiality, addressing security concerns in cloud environments. The prefetching algorithm optimizes retrieval by pre-computing and storing likely retrieval paths based on historical access patterns and feature similarity. By matching query image features with stored video features, the system efficiently retrieves relevant video segments, leveraging the prefetching algorithm to reduce retrieval time and improve performance. This proposed method offers a comprehensive solution for securely storing and efficiently retrieving medical surgical videos, addressing the challenges of scalability, security, and retrieval speed in cloud-based systems.

IV. Experimental Result Analysis

In preprocessing, each video was divided into frames at a rate of 1 frame per second. Key frames were selected based on visual content changes using histogram differences. Features were extracted from the key frames using a Variational Autoencoder (VAE) to reduce dimensionality. Group Lasso was applied to select the most informative features, minimizing noise. Selected features were encrypted using AES and stored in AWS cloud storage. A weighted graph-based prefetching algorithm was used to optimize retrieval paths. Image queries were generated by extracting frames from the dataset. The retrieval performance was evaluated based on speed and accuracy, measured by precision and recall.



UPLOAD DOCUMENTS

View Upload Documents

File Id	File Name	Owner Name	Secret Key	Created Date	Download Encrypted File	Download Decrypted File	Delete Record
7	file	meghala	cmD3up2Ns+CrXO7O	29 Aug 2022 19:05:18 PM	Download Encrypted File	Download Decrypted File	Delete Record
18	cholec surgery 2 tools	meghala	GPcrsW/3k5yy8Tbm	10 Jul 2024 11:58:06 AM	Download Encrypted File	Download Decrypted File	Delete Record
19	cholec surgery 3 tools	meghala	m8M/S/1PevxP2ggj	10 Jul 2024 11:59:51 AM	Download Encrypted File	Download Decrypted File	Delete Record
20	cholec bagging	meghala	Qb5YCyXkZ+4yfCt4	10 Jul 2024 12:02:11 PM	Download Encrypted File	Download Decrypted File	Delete Record
21	cholec surgery triangle dissection	meghala	RK+riFUuSsxbQOgl	10 Jul 2024 12:25:19 PM	Download Encrypted File	Download Decrypted File	Delete Record

LIST OF DOCUMENT DOWNLOADS

List of Download

Request Id	Owner Name	File Name	Public Name	Status	Created Date
3	meghala	file	demo	Download	29 Aug 2022 19:08:36 PM
4	meghala	file	demo	Download	10 Jan 2023 10:42:36 AM
10	meghala	cholec surgery 2 tools	user1	Download	10 Jul 2024 12:06:40 PM
11	meghala	cholec surgery triangle dissection	user1	Download	10 Jul 2024 12:29:52 PM
12	meghala	cholec surgery 3 tools	user2	Download	10 Jul 2024 12:33:45 PM

Fig. 4. Videos uploaded and downloaded.

The proposed framework implementation is done with the hardware requirements of Intel core i5 processor, 16GB RAM, and NVIDIA GeForce GTX 1080 GPU for deep learning model training. Python libraries such as TensorFlow, Keras, Scikit-learn are employed for feature extraction and learning; OpenCV for video processing and PyCrypto for data encryption.

Cholec80 dataset is used for experimentation. The Cholec80 dataset consists of 80 videos of cholecystectomy surgeries, including laparoscopic gallbladder removal procedures. Each video is labeled with tool presence and phase annotations. Videos are provided in a resolution of 1920x1080 pixels. Each video lasts approximately 1 hour, comprising different phases of the surgical procedure.

Fig. 4 illustrates the developed application results of secured video stored and decrypted videos downloaded. Fig. 5 depicts the predicted results of the moment retrieved from the Cholec surgery video for the given image query. The user request to access the video. Only authenticated users are allowed to download the video. Here, the user requests a video using text or image. Those video files matching to the text queries is listed for view access. The image query is passed to the feature selection algorithm and those matching feature set existing video frames are retrieved. This is illustrated in Fig. 5. The average time taken to retrieve the relevant video segment is 0.5 seconds per query. Lower times indicate faster retrieval, which is crucial for real-time applications.

The Table I presents a comparison between existing and proposed methodologies based on four different evaluation metrics: Recall, Intersection over Union (IoU), mean average precision, and ground truth. The existing methodology includes Deep learning methods such as CNN, RCNN, and DNN, while the proposed methodology is the Hybrid approach using VAE,Group Lasso and weighted graph prefetching algorithm. Based on the Recall metric, the proposed Hybrid approach outperforms all existing methods with a Recall of 99.03%, which is higher than the Recall values for all existing methods. Similarly, the proposed methodology outperforms all existing methods for IOU, mean average precision, and ground truth. The accuracy of the proposed methodology is also higher than that of existing methods, with a value of 99%. In contrast, the accuracy values for existing methods range from 93% to 94%. Therefore, the proposed Hybrid approach using PSO shows promising results and can be considered as a potential alternative to the existing Deep learning methods for the given task.

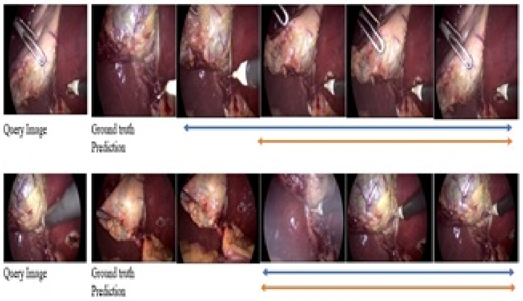


Fig. 5. Query matching video frames.

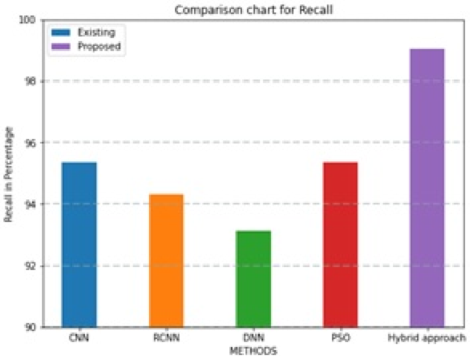


Fig. 6. Comparison of recall.

TABLE I. PERFORMANCE METRICS COMPARISON TABLE

	Recall	IoU	avg precision	Ground truth	Accuracy
CNN	95.36	96.63	93.75	93.98	93%
RNN	94.32	96.21	92.43	92.74	94%
DNN	93.13	95.38	90.86	91.29	93%
PSO	95.36	96.98	93.74	93.97	93%
Proposed approach	99.03	98.40	98.65	99.13	99%

Recall reflects the proportion of relevant video segments that were correctly retrieved out of all relevant segments. Higher recall means fewer relevant segments are missed. Fig. 6 displays recall comparison charts. On this graph, methods are represented by the x axis, and recall expressed as a percentage along the y axis. The high precision and recall indicate that the proposed method accurately retrieves relevant video segments for a given image query, with minimal false positives and negatives.

The IoU comparison is shown in Fig. 7. On the graph,

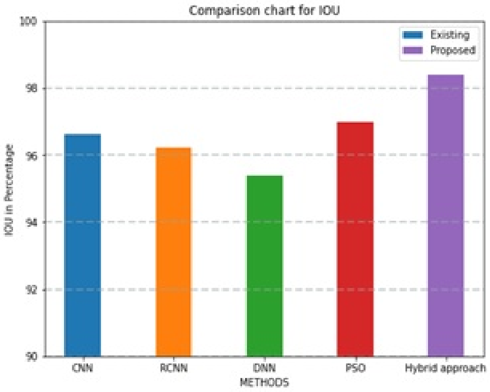


Fig. 7. Comparison of IoU.

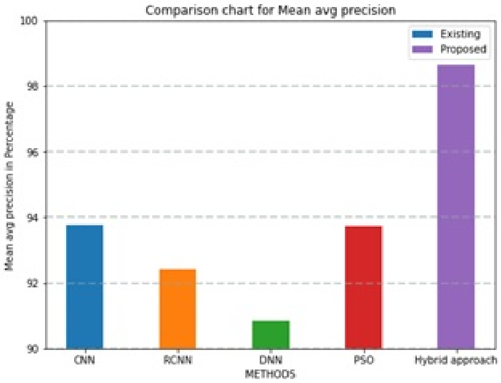


Fig. 8. Comparison of mean avg precision.

methods are represented by the x axis, while IoU is shown as a percentage along the y axis.

Precision indicates the proportion of correctly retrieved video segments that are relevant out of all retrieved segments. Higher precision suggests fewer irrelevant segments are retrieved. The comparison of average precision is shown in Fig. 8.

The ground truth comparison is shown in Fig. 9. On this graph, methods are represented by the x axis, while the percentage of ground truth is shown by the y axis.

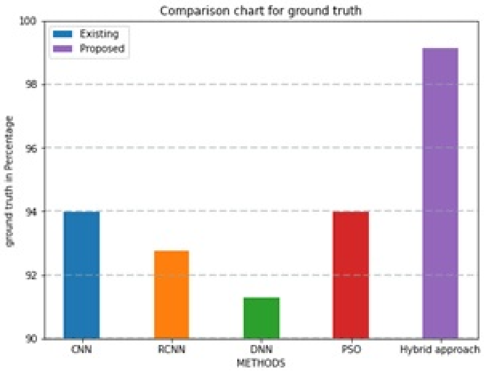


Fig. 9. Comparison of ground truth.

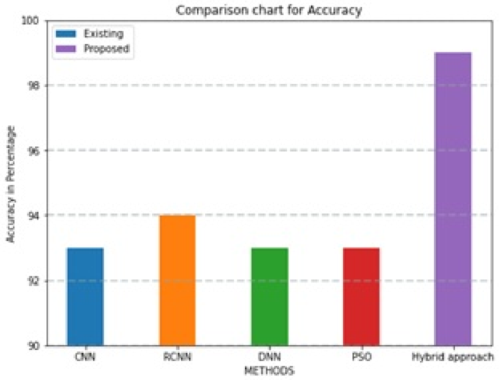


Fig. 10. Performance accuracy.

The accuracy comparison is shown in Fig. 10. The y axis represents the percentage of accuracy, while the x axis displays the various strategies. The proposed framework significantly improves retrieval accuracy and speed compared to traditional methods, making it suitable for real-time applications in medical environments. The hybrid feature selection method effectively reduces noise, enhancing the accuracy of the retrieval process. The cloud-based storage solution provides scalability and flexibility, allowing healthcare institutions to manage large volumes of surgical video data efficiently. Hence the experimental results highlight the effectiveness of the proposed method in optimizing retrieval and secure storage of medical surgical videos using deep learning techniques, offering promising implications for improving healthcare data management.

V. Conclusion

This study presented a novel framework for the secure storage and optimized retrieval of medical surgical videos using a deep learning-based approach, specifically tailored for cloud environments. The framework integrates advanced techniques such as Variational Autoencoder (VAE) and Group Lasso for hybrid feature selection, ensuring that only the most relevant and informative features are retained. These selected features are then securely encrypted using AES and stored in a cloud infrastructure, with retrieval performance further enhanced by a weighted graph-based prefetching algorithm.

The experimental evaluation, conducted on the Cholec80 dataset, demonstrated that the proposed method achieves a high retrieval accuracy, with a precision of 98.65% and a recall of 99.03%, while maintaining a swift average retrieval time of 0.5 seconds per query. These results underscore the effectiveness of the system in not only preserving the security of sensitive medical data but also in providing rapid and accurate access to relevant video segments, which is crucial in medical and surgical contexts.Furthermore, the integration of this framework into cloud platforms like AWS highlights its practical viability, offering a scalable and flexible solution for managing large volumes of medical video data. This approach is particularly valuable in healthcare settings where the secure and efficient handling of video data is essential for both clinical practice and research.

Thus the proposed method provides a robust solution to the challenges of medical video storage and retrieval, contributing significantly to the field by improving data security, retrieval speed, and accuracy. The initial setup and processing require substantial computational resources, which may limit accessibility for institutions with limited hardware capabilities. The integration of deep learning algorithms and encryption techniques can introduce computational complexity, which may affect the performance of the system in environments with limited processing power. The reliance on cloud platforms for storage and retrieval may pose challenges in terms of cost and internet connectivity, particularly for smaller healthcare facilities with limited resources. Future research could incorporate natural language processing techniques to analyze textual annotations with visual features, providing a more comprehensive retrieval system. Further optimization of the retrieval algorithm could enable real-time querying capabilities, essential for immediate decision-making during surgical procedures.

Acknowledgment

The authors would like to thank Vellore Institute of Technology, Vellore for their support.

References

[1] A. I. Al Abbas, J. P. Jung, M. K. Rice, A. H. Zureikat, H. J. Zeh III, and M. E. Hogg, “Methodology for developing an educational and research video library in minimally invasive surgery,” *Journal of Surgical Education*, vol. 76, no. 3, pp. 745–755, 2019.

[2] R. Cao, Z. Tang, C. Liu, and B. Veeravalli, “A scalable multicloud storage architecture for cloud-supported medical internet of things,” *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1641–1654, 2019.

[3] Y. Chen, S. Ding, Z. Xu, H. Zheng, and S. Yang, “Blockchain-based medical records secure storage and medical service framework,” *Journal of medical systems*, vol. 43, pp. 1–9, 2019.

[4] P. S. Deshpande, S. C. Sharma, and S. K. Peddoju, *Security and Data Storage Aspect in Cloud Computing*. Springer, 2019, vol. 52.

[5] A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, “Emerging technologies in data mining and information security,” *Proceedings of IEMIS-2018*, 2018.

[6] K. He, J. Chen, Y. Zhang, R. Du, Y. Xiang, M. M. Hassan, and A. Alelaiwi, “Secure independent-update concise-expression access control for video on demand in cloud,” *Information Sciences*, vol. 387, pp. 75–89, 2017.

[7] H. Li, Y. Yang, Y. Dai, S. Yu, and Y. Xiang, “Achieving secure and efficient dynamic searchable symmetric encryption over medical cloud data,” *IEEE Transactions on Cloud Computing*, vol. 8, no. 2, pp. 484–494, 2017.

[8] A. Lounis, A. Hadjidj, A. Bouabdallah, and Y. Challal, “Healing on the cloud: Secure cloud architecture for medical wireless sensor networks,” *Future Generation Computer Systems*, vol. 55, pp. 266–277, 2016.

[9] Y. Li, K. Gai, L. Qiu, M. Qiu, and H. Zhao, “Intelligent cryptography approach for secure distributed big data storage in cloud computing,” *Information Sciences*, vol. 387, pp. 103–115, 2017.

[10] G. Megala, P. Swarnalatha, S. Prabu, R. Venkatesan, and A. Kaneswaran, “Content-based video retrieval with temporal localization using a deep bimodal fusion approach,” in *Handbook of Research on Deep Learning Techniques for Cloud-Based Industrial IoT*. IGI Global, 2023, pp. 18–28.

[11] D. Pei, X. Guo, and J. Zhang, “A video encryption service based on cloud computing,” in *2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC)*. IEEE, 2017, pp. 167–171.

[12] S. N. Pundkar and N. Shekokar, “Cloud computing security in multiclouds using shamir’s secret sharing scheme,” in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, 2016, pp. 392–395.

[13] P. Srivastava and N. Garg, “Secure and optimized data storage for iot through cloud framework,” in *International Conference on Computing, Communication & Automation*. IEEE, 2015, pp. 720–723.

[14] C. Stergiou, K. E. Psannis, B.-G. Kim, and B. Gupta, “Secure integration of iot and cloud computing,” *Future Generation Computer Systems*, vol. 78, pp. 964–975, 2018.

[15] M. Usman, M. A. Jan, and X. He, “Cryptography-based secure data storage and sharing using hevc and public clouds,” *Information Sciences*, vol. 387, pp. 90–102, 2017.

[16] Priyanka and A. K. Singh, “A survey of image encryption for healthcare applications,” *Evolutionary Intelligence*, vol. 16, no. 3, pp. 801–818, 2023.

[17] G. Megala and P. Swarnalatha, “Efficient high-end video data privacy preservation with integrity verification in cloud storage,” *Computers and Electrical Engineering*, vol. 102, p. 108226, 2022.

[18] M. G. and S. P., “Discrete hyperchaotic s-box generation for selective video frames encryption,” *Journal of Computer Science*, vol. 19, no. 5, pp. 588–598, 2023.

[19] H. Yan, M. Chen, L. Hu, and C. Jia, “Secure video retrieval using image query on an untrusted cloud,” *Applied Soft Computing*, vol. 97, p. 106782, 2020.

[20] Y. Yang, X. Zheng, and C. Tang, “Lightweight distributed secure data management system for health internet of things,” *Journal of Network and Computer Applications*, vol. 89, pp. 26–37, 2017.

[21] G. Megala and P. Swarnalatha, “Stacked collaborative transformer network with contrastive learning for video moment localization,” *Intelligent Data Analysis*, no. Preprint, pp. 1–18.

[22] F. Khelifi, T. Brahimi, J. Han, and X. Li, “Secure and privacy-preserving data sharing in the cloud based on lossless image coding,” *Signal Processing*, vol. 148, pp. 91–101, 2018.

[23] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, “Blockchain for secure ehrs sharing of mobile cloud based e-health systems,” *IEEE access*, vol. 7, pp. 66 792–66 806, 2019.

[24] P. Deshpande, S. C. Sharma, and S. K. Peddoju, “Data storage security in cloud paradigm,” in *Proceedings of Fifth International Conference on Soft Computing for Problem Solving: SocProS 2015, Volume 1*. Springer, 2016, pp. 247–259.

[25] V. Jagadeeswari, V. Subramaniyaswamy, R. t. a. Logesh, and V. Vijayakumar, “A study on medical internet of things and big data in personalized healthcare system,” *Health information science and systems*, vol. 6, no. 1, p. 14, 2018.

[26] M. Kolarik, M. Sarnovsky, J. Paralic, and F. Babic, “Explainability of deep learning models in medical video analysis: a survey,” *PeerJ Computer Science*, vol. 9, p. e1253, 2023.

[27] A. Sánchez-Caballero, D. Fuentes-Jiménez, and C. Losada-Gutiérrez, “Real-time human action recognition using raw depth video-based recurrent neural networks,” *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16 213–16 235, 2023.

[28] V. Biksham and D. Vasumathi, “Homomorphic encryption techniques for securing data in cloud computing: A survey,” *International Journal of Computer Applications*, vol. 975, no. 8887, 2017.

[29] M. M. Islam and Z. A. Bhuiyan, “An integrated scalable framework for cloud and iot based green healthcare system,” *IEEE Access*, vol. 11, pp. 22 266–22 282, 2023.

[30] S. Unar, X. Wang, and C. Zhang, “Visual and textual information fusion using kernel method for content based image retrieval,” *information Fusion*, vol. 44, pp. 176–187, 2018.