JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3730-3743
https://publishoa.com
ISSN: 1309-3452

Swapped Face Detection Using Deep Learning and Subjective Assessment

Shivani Chaudhry[1], Dr. Krishna Kumar Singh[2]
Computer Science & Engineering, STAREX University, Gurugram, Haryana India [1]
Associate Professor, Symbiosis Center for Information Technology, Symbiosis International University, Pune,India[2]

Received:2022 March 15; **Revised**:2022 April 20; **Accepted**:2022 May 10.

Abstract
Face swapping is an exciting new technology that allows the identification of a source face to be transferred to a target face while maintaining the target face's attributes. As a result, this study uses deep learning and subjective assessment for face swapping detection. In this study, face-swapping generates image-realistic and time-coherent sequences. Human recognition can aid in the prevention of face swapping attacks. Human volunteers determine accuracy in reconciling faces. They may believe a photograph is a forgery. Rankings of GIF emotions: A pair of Hamming-LUCBs is employed when you click an image. Selecting a switched type has a 50% failure rate. The CNN-LSTM and RNN-LSTM classifiers are compared. The RNN-LSTM algorithm outperformed the CNN-LSTM algorithm. Rather than that, we learned by exchanging faces. Validation made use of the system's default hyper parameters and epochs. We train and publish separately on both strategies. As a result, we compared them. 25-50-85 LSTM (CNN & RNN). We quantify the model's output. The dimensions of the faces are 0.2865 and 0.0415. R/F equals 0.1106. Real and fictitious faces have values of 0.3701 and 0.4229. Between 0.3741 and 0.1175. In Ae-GAN generated fictitious images, human connectedness is imprecise. The true identities of Nirkin and AEGAN (around 0.09 difference). This study makes extensive use of transfer learning. The Best Face Swap Detection Photos in the World Each model has over 1000 images taken in real life (the largest known). The model was confused because we put it to the test on humans. It makes a comparison of two photos. An image can identify an imposter; the context would then make it very evident that our model does not exist. As a result, the model can perceive a human.

Keywords: Face swapping, New technology, Deep learning, Subjective assessment, Algorithm, Humans, Image

Introduction
Recent research on face modification challenges has resulted in deep network-based techniques (Wöhler et al., 2021). Face swapping transfers a face from a source image to a destination image, smoothly and realistically replacing it. As a result, face-swapping is a widespread issue in computer vision and graphics. Numerous recent proposals have been made regarding automatic face switching. These endeavors have obviated the necessity for tedious manual face editing, hence expediting the advancement of face editing (Lutz & Bassett, 2021).Face swapping is an exciting new technology that allows the identification of a source face to be transferred to a target face while maintaining the target face's attributes. Industrial face swapping is a time-consuming process that involves expensive technology to reconstruct the actor's face model and recreate the lighting in the scene. Recent research has concentrated on face swapping without sophisticated technology (Caruana & Vella, 2020).Source-oriented approaches operate with the source face's image, whereas target-oriented approaches work with the target face's features. The source face is blended into the target image by transferring the target face's attributes (for example, expression and posture) to the source face. Due to their reliance on the stance and lighting of the source image, these techniques cannot precisely duplicate the target's emotional state (Yadav et al., 2020). Target-oriented approaches directly alter the target image's features and are highly adaptable to source face variation. The open-

source method allows for the face-swapping of two distinct identities but not for generalisation. Recent research achieves a high level of realism using a two-stage architecture. However, these techniques overidentify. This is a lax requirement on attribute preservation that frequently leads to mismatches across attributes (Hou, 2018).

Nowadays, face-swapping is a widespread issue in computer vision and graphics. Numerous recent proposals have been made regarding automatic face switching. These endeavours have obviated the necessity for tedious manual face editing, hence expediting the advancement of face editing. Nonetheless, this enabling technology has raised legitimate concerns, most notably its abuse potential. The emergence of "Deepfakes" on the internet has increased public and government concern. As a result, swift countermeasures are required, most notably technological advancements capable of detecting falsified videos (Ma et al., 2021).Digitally edited images and films, notably those created with DeepFake technologies, have caused widespread controversy. "DeepFake" refers to a deep learning system capable of generating fraudulent films by swapping one person's face for another (Radenkovic et al., 2019). Late in 2017, a Reddit user using the handle "deepfakes" claimed to have developed a machine learning system that enabled him to swap celebrity faces into pornographic movies. False news, hoaxes, and financial fraud are just a few of the most harmful applications for this type of bogus information. Researchers are increasingly turning to general media forensics to detect a facial change in photos and videos. The detection of fake faces is based on recent research in anti-spoofing biometrics and data-driven deep learning. In recent years, there has been an increase in the number of workshops at leading conferences, international projects such as MediFor, funded by DARPA (Muluye, 2020).

The study uses deep learning and subjective assessment for face swapping detection. No previous work has employed a subject assessment technique that can be applied to the faces of various subjects without requiring specialised expertise. In this study, face-swapping generates image-realistic and time-coherent sequences. As a result, this study is the first to simultaneously manipulate position, expression, and identification without requiring individual or pair-specific training. Additionally, the current survey complements by delving deeper into each facial alteration group, including manipulation techniques, public databases, and essential benchmarks for technical evaluation of false detection methods, as well as providing a summary of the findings. Additionally, we discuss the current swapped face detection, emphasising its advancements and shortcomings in fake identification.

Related reviews

Face swapping-Muluye (2020) apply a new facial "texture" to the projected 3D model. As a result, estimations include scene variables such as three-dimensional orientation and camera focal length. Comparable to Morphable Modeling, optimizing all model parameters is a similar process. Caruana & Vella (2020) propose replacing the face without 3D reconstruction. The process entails locating a face that is similar to the input face. As a result, a big image library is necessary. A rating algorithm is then used to determine which library image should be substituted. The candidates' lighting and colour settings can be altered to make the swapped face appear more realistic. Their technique can switch between faces that appear genuine to the user. However, it can only be used to swap out one face for another (Ma et al., 2021).Lutz & Bassett (2021) advised that faces in referenced pictures be replaced with similar forms and traits. A technique based on triangulation is utilized to adjust the reference face and background to fit the appearance of the input face to distort the image. The face ROIs are recognised using a parsing algorithm, and the boundaries and colours are adjusted using a Poisson image editing technique. Ashok et al. (2020) provide an overview of the fundamentals of Poisson editing. Once a Laplacian has been generated, the Poisson equation for smooth domain filling can be solved numerically. This technique can also be used to process colour image channels.The practical success of deep learning in image processing has resulted in the creation novel face-swapping techniques. Ansari (2020) employed the style transfer technique. Face and posture are considered instances of content, whereas identity is considered an example of the style. Neural networks are used to convert the image to a new format. Before and after the change, the face is aligned. Face landmarks were recorded in various techniques from the source and target photos, resulting in 3D face models that may be exchanged using transformations. After

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3730-3743
https://publishoa.com
ISSN: 1309-3452
transformation, segmentation and mixing are performed using a fully convolutional neural network (Thenmalar, 2020).With the usage of auto-encoders and generative advertising networks (GANs), face-swapping is becoming increasingly automatic. DeepFake employs both of these tactics. Joint latent space is discovered between the input and output of an auto-input encoder. This untapped space can reconstruct the original image of two or more individuals (Ruelas et al., 2020). Two auto-encoders are trained on the same encoder to learn the latent space. Through this training, the encoder acquires the ability to detect common features of the face (pose and relative expression). Due to the unique nature of each decoder, no one can teach others how to generate realistic representations of a particular individual from latent space. Constructing faces from latent space using the encoder/decoder of the faces of two distinct individuals (Singh, 2021). Sadu & Das (2020) creates the image using an auto-encoder and utilizes a CNN as a discriminator to assess whether the face is genuine or a doppelganger. Empirically, it has been demonstrated that applying this adversarial loss on switched faces improves their quality.
Zhang & Doyle (2020) propose using hair and faces to swap and replace people's faces in the latent space. In this method, an RS-GAN is utilized to build a single face-swapped image (region-separative generative adversarial network). In contrast to image-based face replacement, Wöhler et al. (2021) employ video-based face replacement. They replace faces in videos with low-cost equipment and minimum human participation.

Detecting a forgery--Kedar (2021) used the LFW dataset to create swapped faces. They constructed image features using SURF and BoW rather than raw pixels. They then examined random forests, support vector machines, and simple neural networks. They obtained 92 percent accuracy by focusing exclusively on their swapping methods. Their switched faces are unproven. Additionally, their dataset is small compared to other studies, consisting of only 10,000 images (half swapped).Chandrasekaran (2021) examined the generalisation of the method. They collected 53,000 photographs from 150 recordings in all. The swapped faces in their dataset were generated in a variety of ways. We evaluated phoney face identification using texture-based and CNN-based methods. Smoothing and mixing were used to create a more natural appearance for the switched face.On the other hand, using video frames reduces the variety of sights and enhances their similarity. Agarwal et al. proposed the concept of Weighted Local Magnitude Patterns. They opted for films over photographs. They constructed their own data set. Cui et al. (2020) investigated the detection of video swapped faces. They evaluated many techniques for detecting DeepFakes. Additionally, they evaluate the security of facial recognition systems based on VGG and FaceNet.
Thenmalar (2020) recently assessed a variety of detectors in various environments. They compare their results to human performance on changed images. Our work is strikingly similar to theirs. Rather than utilizing videos, we constructed a massive dataset using still images to solve concerns with picture similarity. Additionally, we provide approximately 1000 different nature photographs of each celebrity taken outdoors. This is advantageous for auto-encoders that require a high number of images to train. Our dataset can detect more than just fraudulent faces using this technique. It is also worth noting that images are evaluated according to their authenticity. By comparing our classifier's score margin ranking to people's, we demonstrate that human and machine certainty are related but not identical (Muluye, 2020).

Methods of Deep Learning--In recent years, deep learning has transformed image classification. AlexNet, a five-layer deep convolutional neural network (CNN), won the ImageNet competition to classify 1.2 million high-resolution pictures. This discovery demonstrated that CNNs outperformed other algorithms for feature extraction, such as SURF and HOG (Ding et al., 2020). The term "generalisation ability" refers to the capacity to classify images with considerable size, shape, colour, or texture variances appropriately. Since deep learning algorithms do not rely on pre-engineered (i.e., manually defined) features to classify images, they exhibit superior generalisation performance. SURF and HOG identified images using user-supplied features (e.g., edges, corners, and gradients) (Kedar, 2021). On the other hand, deep learning is based on feature learning, which does not require pre-defined features. Rather than that, a system called back-propagation automatically determines the most useful attributes for photo identification. Among these qualities are edges, corners, and gradients. As a result, systems based on deep learning can detect intricate patterns in images, exceeding previous methods such as SURF and HOG (Ashok et al., 2020).

Methods for detecting fake face---The synthesis, editing, alteration, and transfer of facial imagery are extensive studies. This study will conduct a brief assessment of existing techniques.

Encoder-Decoder Methods--Zhang et al. (2021) provide an encoder-decoder model that has implications for the current work. Coupled weights at multiple layers of the encoders and decoders closest to the encoded bottleneck establish a common latent space for encoded pictures. Hou (2018) discuss face switching as difficulty in style transmission in KSDT17. This network can be used to quantify the content and style loss in the VGG-19 feature space. Ma et al. (2021) examine our model structure's small sample. They can instruct the decoders to reconstruct the originals by bending the input images. Korshunova et al. (2017) demonstrated a multitasking encoder-decoder system's remarkable face switching performance. Compared to our unsupervised approach, their model necessitates many labelled training data. Maity et al. (2021) employ a separate generator to merge the target face with the source image.

GAN Techniques--Since Cui et al. (2020) created GANs (generative adversarial networks) for image synthesis, GANs have grown popular. The most successful method for face swapping is to use GAN-based image-to-image translation. Matched data must be collected, which might be challenging to make this strategy work. There are several methods for easing or eliminating this constraint (Muluye, 2020).
Combining one face-hair separator network with another hair-face network results in a GAN that "checks" and "corrects" the findings. The researchers employed a GAN to extract information on the surface normal, albedo, lighting, and alpha matte of input photographs to make more interesting images. Anwer et al. (2021) use the Face Action Coding System to synthesize facial expressions from a single photo. Cerda et al. (2021) recently produced GAN facial animation. Birkin et al. devised a face-swapping and reenactment pipeline capable of generalizing new faces. This method produces slightly blurred images, which are unsuitable for high-quality work.

Geometric Methods (Morphable Models) --Geometric Methods refer to parametric models of the human face. The vector space of morphable three-dimensional models is defined by learning exemplars from photographs and three-dimensional scans. Recent work has increased the capability of deep encoder-decoder networks to train such models from 2D images. Morphable models are distinct from detailed geometric models capable of capturing individual faces in this context (Shekar et al., 2020). Hou (2018) swapped faces using morphable models; however, the results were not photorealistic. Lutz & Bassett (2021)employed an edit-based technique with additional post-processing to improve the match between source and target photos. Maity et al. (2021) transfer specific facial components using a geometric method, whereas Bhavya et al. (2021) specialises in photo editing to remove closed eyes and look-aways. Chen et al. (2019) generates three-dimensional models of humans from single frontal two-dimensional photos using colour transfer and multi-resolution spline algorithms. Anwer et al. (2021) describe a face-swapping method that uses semi-supervised data and three-dimensional models to register points for image intensities.

Reenactment and Puppeteering--Face swapping is distinct from the problem of face recreation. In contrast to the first scenario, in which the source's behaviour is mirrored on the target's face, the identification is retained. Face swapping reverses the situation: the source's identity is transferred to the target's appearance. However, a recent study indicates that human observers can detect face swaps, highlighting the problems inherent in our existing technique (Zhang & Doyle, 2020).

Pipeline of Face-Swapping--This part demonstrates our robust approach for creating lifelike face swaps at megapixel resolution. The model is the most critical pipeline component, which we will examine next. Additionally, the method for keeping temporal consistency in the swapped photographs is discussed. Then we present our strategy for retaining light and contrast during the compositing process (Cárdenas et al., 2019).

Network Architecture and Training-Domain transfer is used to transform an individual's identity. All identities' images are encoded in a shared latent space and then decoded back to pixel space using the decoder for the desired source appearance. This paper extends the concept of domain transfers and face-swapping to P arbitrary domains (Cárdenas et al., 2019). As a result, the network's encoding path is unbranched, while its decoding path is. This architecture is a comb model, with decoders acting as the "teeth." Our ablation experiment demonstrates that training the model with multiple identities improves the fidelity of expressions in random order subsets. The data is randomly translated, rotated, and scaled. Only the reconstruction loss associated with the currently evaluated facial identity is minimized (Favole et al., 2020).

Additionally, we do not impose exchanging or cycling. We multiply the input and output by mxp to see only the inside of the face. Thus, we minimise a level-dependent loss function compared to a two-way model. Additionally, the outputs of the multi-way decoder can correspond to many identities or the same identity in different illumination conditions. By utilising a single network, training time is reduced compared to developing all possible appearance pairs using two-way networks (Oh, 2020). Our network is educated in a non-adversarial environment through a progressive regime. The network's performance improves when trained on higher-resolution images obtained by down sampling high-resolution input data. The appendix discusses progressive training in further detail. The resolution of the training input determines the network's output. In the absence of high-resolution training data, super-resolution approaches should be viewed as a pre-or post-processing step that enhances the model's output. Because most super-resolution methods for faces include task-specific priors, they are likely to outperform general-purpose algorithms (Singh, 2021).

We partition the data X into P subsets corresponding to a distinct identity. All available samples are scaled to 1024 1024. The early photographs in the progressive regime will be scaled down to 1024 1024, while the final image will be scaled up to 1024 1024. The facial image is aligned to the average landmark locations at the required resolution using an affine transformation. Our implementation is based on the standard 68 landmark point set (Verdhan, 2021). Each normalized facial image is assigned a binary mask mxp. This mask is defined by the convex hull of the 68 standard face markers on XP. A 10% increase in the mask size ensures that critical features such as the brows are not lost owing to slight landmark misalignment. Within the convex hull, the values are 1, while outside the convex hull, they are 0 (Verdhan, 2021).

Alignment and Stability of Landmarks-These tactics are intended to increase the precision of publicly available benchmarks, often single images. While there are specific video benchmarks, their results are rarely tested for temporal consistency. While this is not a concern in many situations, it dramatically affects the realism of the exchanges in our task (Kumar, 2020). Face normalisation is based on face landmarks, which means that slight changes in network inputs result in minor network outputs. When working with high-resolution data, minute differences between frames are amplified, resulting in temporally inconsistent results. This results in facial trembling and distortion. This issue may be resolved by increasing the resolution of a face feature localisation algorithm. However, because most existing data sets are insufficient for this task, trembling may persist (Yadav et al., 2020).

Rather than that, we propose stabilising existing landmark-localisation approaches to address the challenges associated with high-resolution sequential data. The process begins with initial detection and alignment. As a starting point, we take the width of the face bounding box. The following step n times re-initializes the bounding box in w different directions of the image plane (Li et al., 2020). The average of the localized landmark points is obtained after each translation. This strategy effectively minimises variance by mitigating the influence of high resolution on the uncertainty associated with landmark placement. In this case, it entails aggregating and averaging n landmark estimates. Using n = 9 and 0.05, we eliminated all detectable temporal distortions at a resolution of 1024 1024 (Cárdenas et al., 2019).

Compositing with Contrast in Multiple Bands--Even when the source and target images are completely aligned in position and facial expression, composing a source face onto a target image remains challenging. When copying a source onto a target, photometric misalignment creates visible seams. Numerous techniques allow Poisson blending to provide continuous cloning in the gradient domain. However, if the source and target faces are illuminated differently, this may result in visible artefacts in the interior of the face (Shekar et al., 2020).
Edri et al. (2021) used multi-band blending as a new approach to Poisson mixing. A mask specifies the source image area to be cloned into the destination image. The two images are decomposed into a Laplacian pyramid, and the transitions between them are smoothed at each level near the mask's boundaries. However, the cloned region's illumination will be incorrect in our situation (Sadu & Das, 2020). This is why we recreate the coarsest (low-frequency) layers of the target's Laplacian pyramid and blend just the remaining detailed levels—reconstructing the final image from the Laplacian pyramid that has been merged. Additionally, we demand that the boundary smoothing effect extends just into the interior of the face. As a result, the blending does not eliminate the external facial contour (Ruelas et al., 2020).Our enhanced multiband blending method can produce surprising compositing results when the source and target are collected in disparate environments. On the other side, multiple band mixing is incapable of compensating for differences in source-to-target contrast (Kodali & Rekha, 2021).
This way, the contrast between the created source face and the target face can be matched. This can be accomplished by utilizing the Global Contrast Factor (GFC). GFC is used to weigh the local contrast levels at each image scale. To find the contrast coefficient, subtract the GFC of the network output from the GFC of the target image. Finally, this coefficient is multiplied by each pixel. As a result, high-quality compositing is created independent of the capture conditions. In our ablation experiment, we conducted a great deal of comparison (Cerda et al., 2021). The blending mask is carefully selected to prevent the network's edges from introducing the cloned face. Rather than completely encircling the face, we reduce the convex hull of the mask, which is defined by the outer facial landmarks (Bhavya et al., 2021).

Challenges--Open-source face-swapping software and apps generate more Deepfake videos and exert a greater influence on social media. Recognising and filtering such video content has become a technical challenge. The primary impediment to creating a Deepfake forgery detection technique is the lack of high-quality Deepfake and original video datasets suitable for research purposes as training datasets. The detection algorithms and associated packages are incompatible with conventional human user interfaces. These concerns tend to stymie research (Li et al., 2020).

Dataset

Face swapping via auto-encoding necessitates numerous photographs of the same person (usually several hundred). Academic researchers can request access to this dataset's version 1.0 at link 1.
Our celebrity images are obtained using Google's picture API. We execute routines on these images when they are received to delete blank images and duplicates. Then we crop to eliminate superfluous backdrops (Prashanth et al., 2019). Cropping was automated, and visual consistency was evaluated. Our approach uses face and light detection to

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3730-3743
https://publishoa.com
ISSN: 1309-3452
use images with a broader, more diverse range of backdrops. Due to the increased sensitivity of the Auto-Encoder-GAN (AE-GAN) Ding et al. (2020) approach to the background, we eliminate as much as feasible. In practice, a face identification algorithm would be used to determine a region of interest before switching inside it—neural networks with convolutional layers. We select celebrities according to their gender and skin tone.

Method

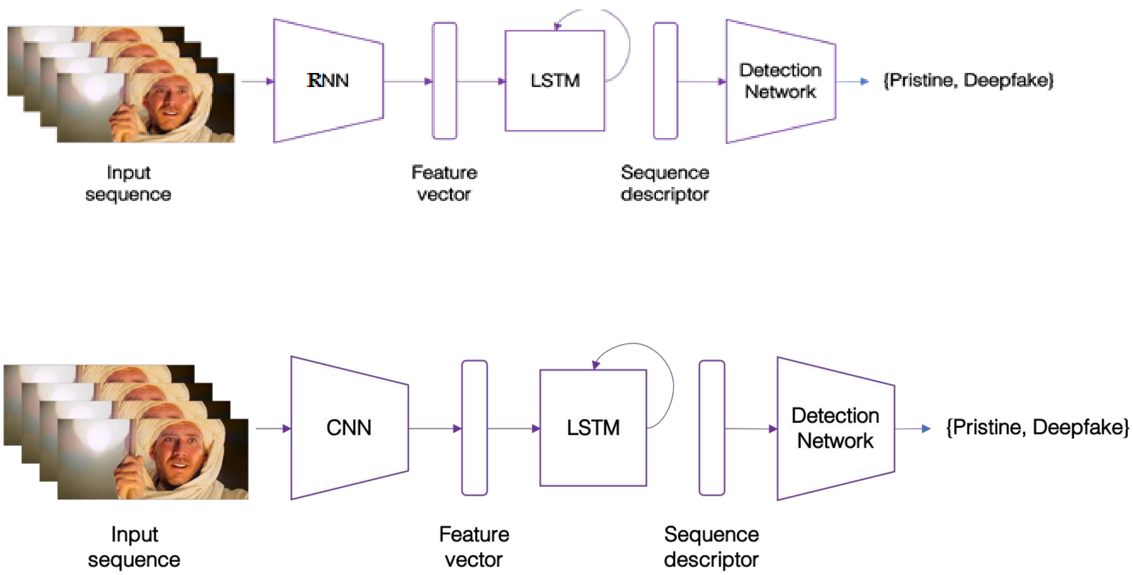

Figure 1: Proposed flowchart

Sub-Subject Human

Face swapping attacks attempt to deceive onlookers, making it vital to understand how humans recognise swapped faces. The purpose of this study is to rank photographs from most authentic to most fraudulent, depending on the accuracy with which human volunteers recognise switched faces. The belief that an image is fabricated may be unambiguous, or the rater may be uncertain. We argue that it is critical to model uncertainty. If the fabricated images are visually unrecognizable, the human evaluation should also be near what the machine learning method predicted (Sadu & Das, 2020). We need to elicit several ratings for each image pair from various raters to gather valid ratings. Even a few photographs are impractical. As a result, we employ two ways to alleviate the ranking stress. As a result, we rate only a subset of the overall photos. We manually picked 100 high-quality swapped faces and 200 real faces from each approach. This is because faces that have been haphazardly replaced are instantly recognised. An attacker would almost certainly "re-select" only the best images before using them maliciously. Active learning is used in the second technique to dynamically select the next image pair to compare for approximate ranking (Ding et al., 2020).

Evaluations of websites--Our website was inspired by the GIFGIF initiative, allowing users to rank GIF emotions. When the evaluator clicks on either image, a box appears. The evaluator may choose to use a registered account or not. Two Hamming-LUCB instances were employed for the face-swapping methods AE-GAN and Nirkin's approach. Attempting to select one of the swapped kinds has a 50% chance of failing. We recruited the assistance of volunteers to rate the images throughout three months. A lecture and an example rating are supplied to ensure that new raters understand the selection procedure (Kedar, 2021).

Results and Discussion

We separated an individual's training and testing sets to evaluate our classifier's performance for CNN-LSTM and RNN-LSTM in Figure 2. RNN-LSTM showed a high percentage compared to CNN-LSTM (Table 1). The system recognised generalising artefacts within swapped face pictures rather than learning specific facial traits. We used default hyper-parameters and a fixed number of epochs for models, eliminating a separate validation dataset. We integrate the switched faces from both techniques during training but report prediction results separately. We collected different frames comparisons from a different model (Table 2). Multiple frames comparisons were Frame-25, Frame-50, Frame-85 for both CNN-LSTM and RNN-LSTM, respectively (Figure 3 to 5). We compare our results quantitatively for the models. In practice, quantifying the results of the face-swapping technique is challenging. Pose and expression provide critical information regarding the outcome of the face swap. SSIM is also used to compare the target and swapped subjects' structural similarity and perceptual loss (Wöhler et al., 2021).

Three different user studies are conducted to evaluate the suggested paradigm. For instance, users can select the most comparable identity, head position and facial expression, or the most realistic. Each study unit includes data from FaceSwap, Zhang et al. (2021), DeepFakes, and our own reshuffled face-swapping. Select a single face that most closely matches our description. Additionally, we compared the Training Accuracy, Validation Accuracy, and Testing Accuracy for both models for each frame. The suggested techniques were compared to one another following (Patidar & Bains, 2021).

Table 1: Comparison between CNN-LSTM and RNN-LSTM

Model	Training Accuracy %	Validation Accuracy %	Testing Accuracy %
CNN-LSTM	99.5	96.9	96
RNN-LSTM	99.65	97.8	98

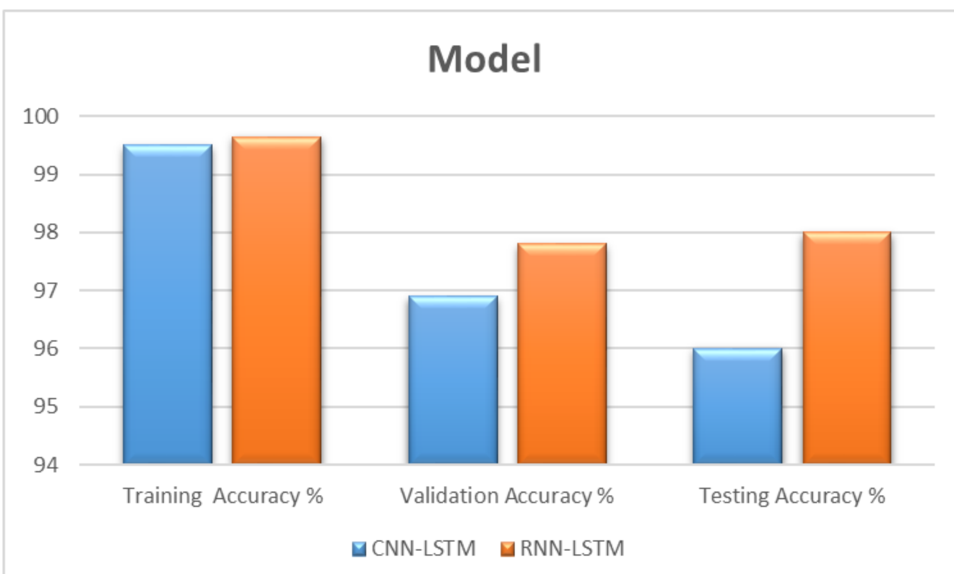

Figure 2: Comparison between CNN-LSTM and RNN-LSTM

Face Swaps are obvious; several indicators of evident face swap artefacts and features were discovered during our tests. Surprisingly, only a small percentage of high-grade alterations reported contour issues due to mismatched

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3730-3743
https://publishoa.com
ISSN: 1309-3452
facial forms. In comparison, ten out of twenty participants observed manipulations of low quality (Yadav et al., 2020). The blurring of the face was frequently attributed to additional changes such as beauty filters. Rather than obvious artefacts, participants frequently remarked unusual eye and mouth movements when performing high-grade face swaps. Some even asserted that the performers' eyes were lifeless or blind. These findings show that viewers can disprove high-grade modifications by examining behavioural indicators such as artificial expressions (Chen et al., 2019).

Table 2 Comparison Different Frame

Model	Training Accuracy %	Validation Accuracy %	Testing Accuracy %
CNN-LSTM Frame-25	99.4	96.95	96.78
RNN-LSTM Frame-25	99.7	97.7	97.8
CNN-LSTM Frame-50	99.5	97	97.1
RNN-LSTM Frame-50	99.65	97.65	97.85
CNN-LSTM Frame-85	99.6	97.5	97.6
RNN-LSTM Frame-85	99.75	98.5	98.65

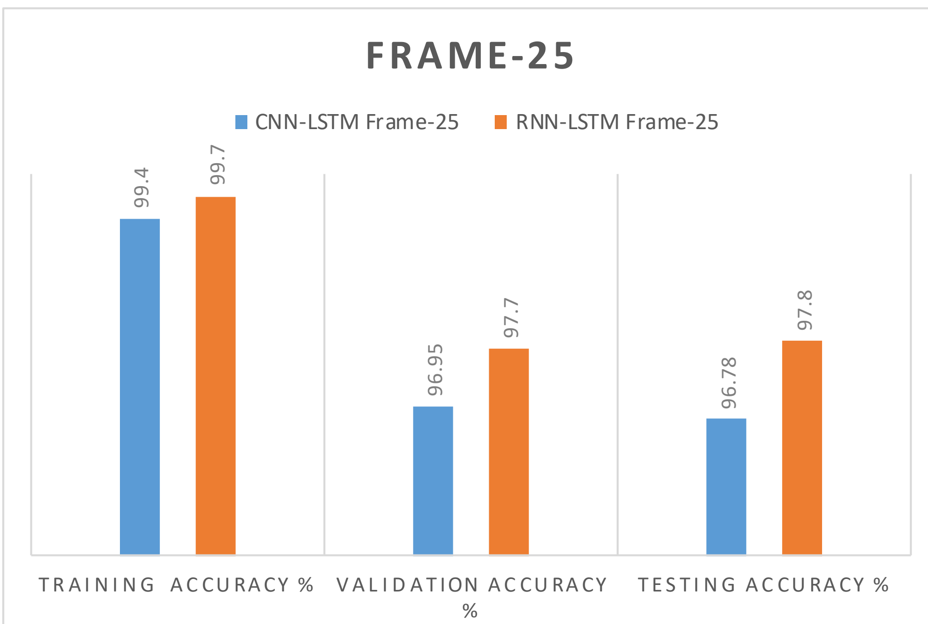

Figure 3: Comparison Frame-25 for between CNN-LSTM and RNN-LSTM

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3730-3743
https://publishoa.com
ISSN: 1309-3452

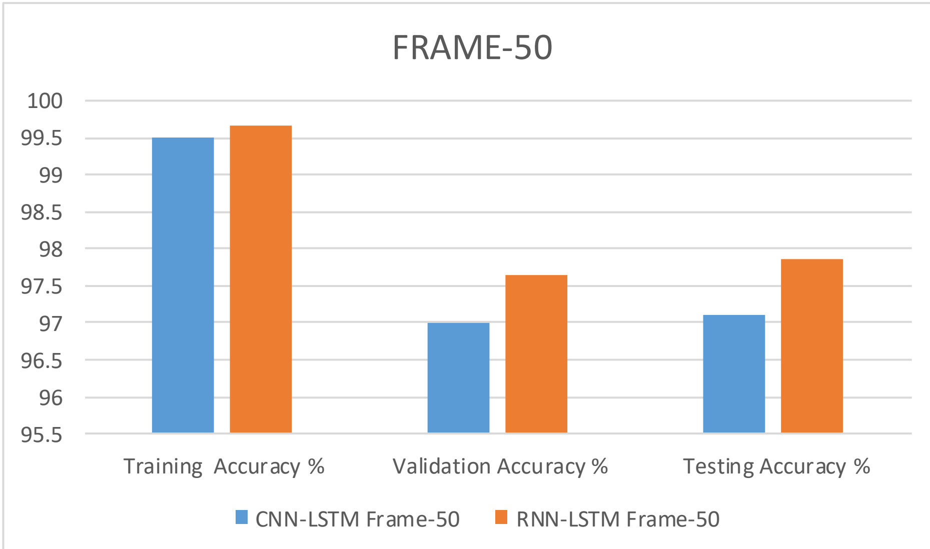

Figure 4: Comparison Frame-50 for between CNN-LSTM and RNN-LSTM

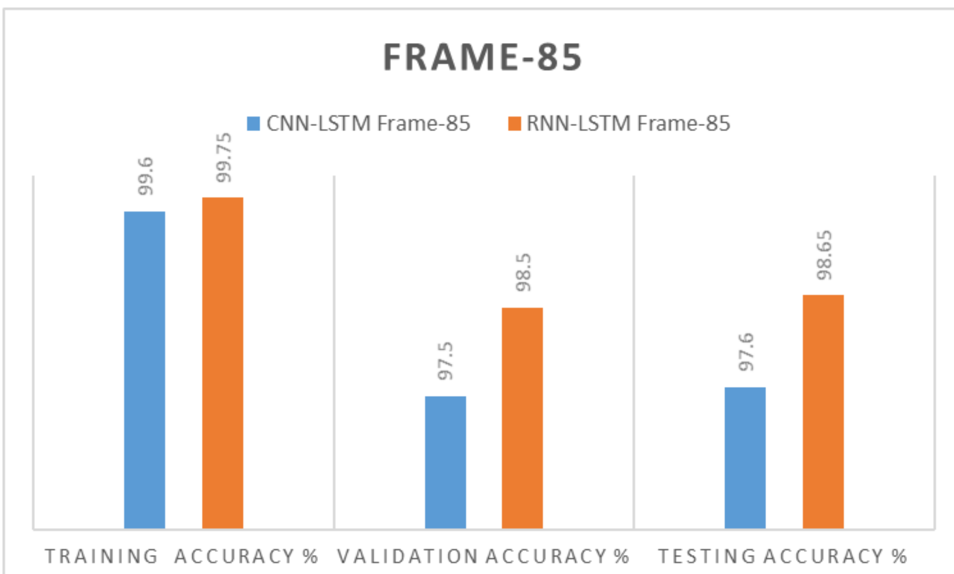

Figure 5: Comparison Frame-85 for between CNN-LSTM and RNN-LSTM

Rather than comparing the accuracy of detecting swapped faces in human individuals, we compare the ranks. Is the ResNet model capable of rating photos that are tough for humans to rate? Or is the model's ranking truly distinct from that of a human?

Analyse the neural network's outputs and compare them to the human ranking. The model learns a convolutional representation of the data during training. In a high-dimensional space, class instances are regularly pushed apart. However, the separation of two occurrences is not always interpretable. Despite this, the output of the activation function may be interpreted as a relative probability of class membership (Lutz & Bassett, 2021).

Predicting whether an instance is genuine or not is believed to be more confident in the last fully linked layer (prior to softmax activation) (i.e., a measure of certainty). Figure 6 shows the score margins of our model and humans for

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3730-3743
https://publishoa.com
ISSN: 1309-3452

the 200 faces used in each technique. Pearson correlation coefficients between scores are 0.7896, and Spearman rank-order correlation coefficients are 0.7579 for this strategy. The AE-GAN Method has correlation values of 0.8332 and 0.7576. On the other hand, the general link may suggest that humans and classifiers are well separated (Sadu & Das, 2020).

Additionally, we present each class's correlations to determine whether they relate to human perception of "reality." Pearson correlation coefficients for natural and artificial faces are 0.2865 and 0.0415, respectively. The Spearman rank-order correlation between real and fake faces is 0.1106. Pearson correlation coefficients for actual and fake faces are 0.3701 and 0.4229, respectively. It is 0.1175 for real faces and 0.3741 for made-up ones. Our model's certainty level and the ratings of human participants are related, but not perfectly—particularly for phoney photographs created using the AE-GAN approach, where there is little human link. Similarly, the correlations between the true faces for Nirkin's and AEGAN approaches are quite close (around 0.09 difference) (Verdhan, 2021). The linear connection is resilient to outliers because it uses just about 50 data points per calculation. As a result, these relationships are inconclusive. As a result, Nirkin's correlation is positive, indicating that the model learns a binary threshold and a rating of the images from most fake to most real. This research indicates that human reality is more easily replicated than "fakeness." We anticipate future efforts to strengthen this closeness in ranking (Zhang et al., 2021).

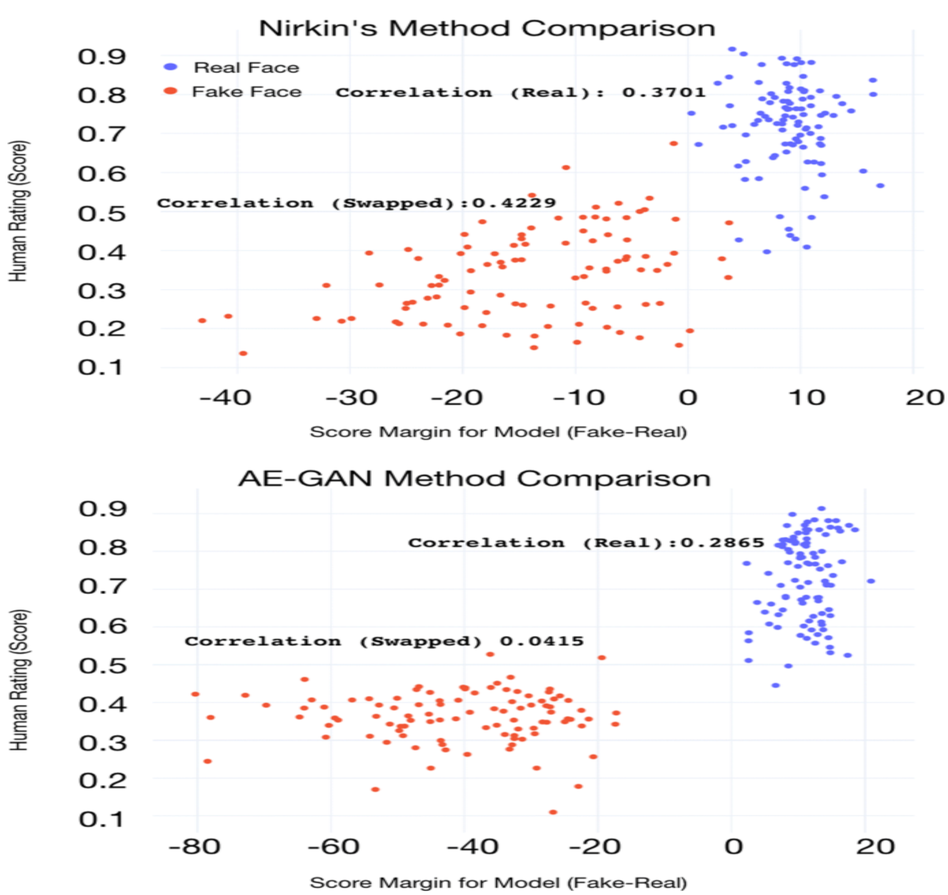

Figure 6: Nirkin's Method Pearson's correlation coefficient is 0.371 in both cases. AE-GAN Method 0.2865 for the real face, 0.0415 for the flipped face

Conclusion

The use of deep transfer learning enabled the detection of swapped faces in this investigation. We created the world's largest face-swapping detection dataset using still images. Face-swapping models trained with AE-GAN benefit from a dataset that contains over 1000 authentic photos for each person (the largest known). In this investigation, we discovered that the model could detect switched faces. Additionally, we compare the model's performance to that of actual humans. A gateway has been created to collect pairwise image comparisons from our dataset. As a result of these comparisons, an estimated ranking is established. In comparison to humans, our deep learning model scores well. By sharing our code, the research community will use our model as a foundation for developing improved deepfake algorithms. This research could aid in developing and evaluating future image forensic methods.

Finally, the models are examined for "fakeness" rather than "identity theft" or "misrepresentation of identity." These two concerns are inextricably linked. That is, our model is capable of recognising swapped faces, the first step toward stealing someone's identity. We appreciate that individuals may identify a fabricated identity even if the photograph appears legitimate. Our model would be unnecessary in that case, as the context surrounding the subject would enable a reasonable person to recognise their absence. As a result, the proposed model can work with a human to recognise an image.

References

1. Ansari, N. (2020). *Dual Shot Face Detecting using Deep Learning*.
2. Anwer, M., Ahmed, G., Akhunzada, A., & Siddiqui, S. (2021). *Intrusion Detection Using Deep Learning*. 1–6. https://doi.org/10.1109/ICECCME52200.2021.9590852
3. Ashok, P., Vashisht, P., Kong, H.-O., Witt-Doerring, Y., Chu, J., Yan, Z., Oort, E. van, & Behounek, M. (2020). *Drill Bit Damage Assessment Using Image Analysis and Deep Learning as an Alternative to Traditional IADC Dull Grading*. https://doi.org/10.2118/201664-ms
4. Bhavya, P., Sharmila, C., Sadhvi, Y., Prasanna, C., & Ganesan, V. (2021). *Pothole Detection Using Deep Learning* (pp. 233–243). https://doi.org/10.1007/978-981-16-1773-7_19
5. Cárdenas, R., Beltrán, C., & Gutiérrez, J. (2019). Small Face Detection Using Deep Learning on Surveillance Videos. *International Journal of Machine Learning and Computing*, *9*, 189–194. https://doi.org/10.18178/ijmlc.2019.9.2.785
6. Caruana, M., & Vella, J. G. (2020). *3D Facial Reconstruction from 2D Portrait Imagery*. https://doi.org/10.11610/isij.4724
7. Cerda, B., Yuan, S., & Chen, L. (2021). *Phishing Detection using Deep Learning* (pp. 117–128). https://doi.org/10.1007/978-3-030-71017-0_9
8. Chandrasekaran, A. (2021). Drowsy Face Detection using Deep Learning Algorithms. *International Journal for Research in Applied Science and Engineering Technology*, *9*, 2174–2179. https://doi.org/10.22214/ijraset.2021.35526
9. Chen, D., Chen, Q., Wu, J., Yu, X., & Jia, T. (2019). Face Swapping: Realistic Image Synthesis Based on Facial Landmarks Alignment. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2019/8902701
10. Cui, S., Zhou, Y., Wang, Y., & Zhai, L. (2020). Fish Detection Using Deep Learning. *Applied Computational Intelligence and Soft Computing*, *2020*, 1–13. https://doi.org/10.1155/2020/3738108
11. Ding, X., Raziei, Z., Larson, E. C., Olinick, E. V., Krueger, P., & Hahsler, M. (2020). Swapped face detection using deep learning and subjective assessment. *EURASIP Journal on Information Security*, *2020*(1), 6. https://doi.org/10.1186/s13635-020-00109-8
12. Edri, E., Armon, N., Greenberg, E., Moshe-Tsurel, S., Lubotzky, D., Salzillo, T., Perelshtein, I., Tkachev, M., Girshevitz, O., & Shpaisman, H. (2021). Laser Printing of Multilayered Alternately Conducting and Insulating Microstructures. *ACS Applied Materials & Interfaces*, *13*(30), 36416–36425. https://doi.org/10.1021/acsami.1c06204
13. Favole, F., Trocan, M., & Yilmaz, E. (2020, November 30). *Melanoma Detection Using Deep Learning*. https://doi.org/10.1007/978-3-030-63007-2_64
14. Hou, X. (2018). An investigation of deep learning for image processing applications. *Undefined*. https://www.semanticscholar.org/paper/An-investigation-of-deep-learning-for-image-Hou/ebc1219ead03fe5720b1d118a5733f6e750ae9fe
15. Kedar, E. (2021). Sarcasm Detection using Deep Learning. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*, 57–64. https://doi.org/10.17762/turcomat.v12i1S.1558

16. Kodali, R., & Rekha, D. (2021). *Face Mask Detection Using Deep Learning*. 1–5. https://doi.org/10.1109/ICCCI50826.2021.9402670
17. Korshunova, I., Shi, W., Dambre, J., & Theis, L. (2017). *Fast Face-Swap Using Convolutional Neural Networks*. 3697–3705. https://doi.org/10.1109/ICCV.2017.397
18. Kumar, S. S. (2020). *Leveraging Big Data and Deep Learning for Economical Condition Assessment of Wastewater Pipelines*. https://doi.org/10.25394/PGS.12218432.V1
19. Li, X., Lang, Y., Chen, Y., Mao, X., He, Y., Wang, S., Xue, H., & Lu, Q. (2020). Sharp Multiple Instance Learning for DeepFake Video Detection. *Proceedings of the 28th ACM International Conference on Multimedia*, 1864–1872. https://doi.org/10.1145/3394171.3414034
20. Lutz, K., & Bassett, R. (2021). DeepFake Detection with Inconsistent Head Poses: Reproducibility and Analysis. *ArXiv*.
21. Ma, T., Li, D., Wang, W., & Dong, J. (2021). *CFA-Net: Controllable Face Anonymization Network with Identity Representation Manipulation*.
22. Maity, S., Das, P., Jha, K. K., & Sekhar Dutta, H. (2021). *Face Mask Detection Using Deep Learning* (pp. 495–509). https://doi.org/10.1007/978-981-16-3067-5_37
23. Muluye, W. (2020). Deep Learning Algorithms and Frameworks for Deepfake Image and Video Detection: A Review. *International Journal of Engineering and Computer Science*, *9*(`10), 25199–25207. https://doi.org/10.18535/ijecs/v9i`10.4533
24. Oh, B.-W. (2020). Map Detection using Deep Learning. *JOURNAL OF ADVANCED INFORMATION TECHNOLOGY AND CONVERGENCE*, *10*, 61–72. https://doi.org/10.14801/JAITC.2020.10.2.61
25. Patidar, S., & Bains, I. (2021). *Intrusion Detection Using Deep Learning* (pp. 113–125). https://doi.org/10.1007/978-981-33-4305-4_10
26. Prashanth, P., Vivek, K., Reddy, D., & Aruna, K. (2019). *Book Detection Using Deep Learning*. 1167–1169. https://doi.org/10.1109/ICCMC.2019.8819725
27. Radenkovic, D., Keogh, S. B., & Maruthappu, M. (2019). Data science in modern evidence-based medicine. *Journal of the Royal Society of Medicine*. https://doi.org/10.1177/0141076819871055
28. Ruelas, B., Dang, H., Nguyen, M., Dao, T., & Dhamo, D. (2020). *Truck Detection Using Deep Learning*.
29. Sadu, C., & Das, P. (2020). Swapping Face Images Based on Augmented Facial Landmarks and Its Detection. *2020 IEEE REGION 10 CONFERENCE (TENCON)*. https://doi.org/10.1109/TENCON50793.2020.9293884
30. Shekar, G., Revathy, S., & Goud, E. (2020). *Malaria Detection using Deep Learning*. 746–750. https://doi.org/10.1109/ICOEI48184.2020.9143023
31. Singh, S. (2021). *Pneumonia Detection using Deep Learning*. 1–6. https://doi.org/10.1109/ICNTE51185.2021.9487731
32. Thenmalar, R. (2020). Intrusion Detection using Deep Learning. *International Journal for Research in Applied Science and Engineering Technology*, *8*, 628–631. https://doi.org/10.22214/ijraset.2020.31521
33. Verdhan, V. (2021). *Object Detection Using Deep Learning* (pp. 141–185). https://doi.org/10.1007/978-1-4842-6616-8_5
34. Wöhler, L., Castillo, S., Zembaty, M., & Magnor, M. (2021). Towards Understanding Perceptual Differences between Genuine and Face-Swapped Videos. *CHI*. https://doi.org/10.1145/3411764.3445627
35. Yadav, D., Renu, Ankita, & Anjum, I. (2020). *Accident Detection Using Deep Learning*. 232–235. https://doi.org/10.1109/ICACCCN51052.2020.9362808
36. Zhang, D., & Doyle, D. (2020). *Gate Detection Using Deep Learning*. 1–11. https://doi.org/10.1109/AERO47225.2020.9172619
37. Zhang, H., Ben, Y., Zhang, W., Chen, T., Yu, G., & Fu, B. (2021). Fine-grained Identity Preserving Landmark Synthesis for Face Reenactment. *ArXiv*.

Author Profile

Dr. Krishna Kumar Singh is an educator, research analyst, writer and trainer with more than 20 years of experience. He is certified professional of SAP S/4HANA, SAP Cloud analytics, SAP ERPsim, SAS analytics, Salesforce, Cortex games, Google and big data analytics. He is an associate professor at Symbiosis centre for information technology, Symbiosis International University, Pune and providing I.T training to personnel of different Industrial/corporate houses like Steel authority of India Ltd, Infosys, TCS, Oriental Bank, Axis Bank (Axis Bank Young Banker's program), Indian army, Navy, air force and many more. He 24 has provided training to the faculty members during different FDP's (faculty development program) on data science at Lebanese French University, Iraq, Erbil, AKGIM Ghaziabad, U.P.College Varanasi, AGBS Pune, Graphic era University, IEC group of institutions Greater Noida, GNOIT Greater Noida, NMIMS Mumbai, Amity University Noida and many more. He

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3730-3743
https://publishoa.com
ISSN: 1309-3452

has associated with National Institute for Entrepreneurship for Small Business Development (NIESBUD) a ministry of Skill initiative as trainer in the area of data analytics and conducted various training programs in information technology. He is also associated with various prestigious B-Schools in Delhi, NCR as guest or visiting faculty AIMA, ICFAI and Sharda University to name a few. Previously he was associated with Amity University, Noida, Graphic era Institute of Technology, Dehradun, College of engineering IILM academy and Galgotias College of engineering and technology as a faculty member. He is an alumnus of Banaras Hindu University and G. B. Pant engineering college, Uttarakhand. He had cleared GATE examination twice. Dr. Krishna had received Scholarship from World Bank (under TEQIP) for Ph.D and from Ministry of HRD for M.Tech. His area of research is machine learning, business analytics, financial BIG data analytics, cloud computing and green computing. There are 06 students pursuing Ph.d under his guidance. There are 05 patent, more than 30 research papers and more than 65 citations in SCI or SCOPUS indexed journals/conferences to his credit. He has written six books on Big data analytics, fraud detection analytics, MASEEI (management in action social, economic and ethical issues), Cloud computing, Mobile marketing and financial market analytics. He is an editorial board member of different SCOPUS indexed research journal and IEEE conferences. He can be contacted at singhkrishna80@gmail.com, krishnakumar@scit.edu, +91-9810866492.

Shivani Chaudhry is pursuing PHD (Computer Science & Engineering) from STAREX University , Gurugram, Haryana. She has completed her M.tech (Computer Science & Engineering) from Deenbandhu Chhotu Ram University of Science &Technology (DCRUST), Murthal ,Sonipat, Haryana, India in 2018 and She has completed her B.tech (Computer Science & Engineering) from SITM, Sonipat, India in 2016.