A New Hybrid CNN-LSTM Model with Non-SoftMax Functions for Face Spoof Detection

S Lokesh Kumar
VIT University - Chennai Campus
Yamani Sai Asish
VIT University - Chennai Campus
Sannasi Ganapathy (sganapathy@vit.ac.in)
VIT University - Chennai Campus https://orcid.org/0000-0001-9177-5378

Research Article

Keywords: Classification, deep learning, CNN, LSTM, space spoof detection, face recognition

Posted Date: August 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-837209/v1

A New Hybrid CNN-LSTM Model with Non-SoftMax Functions for Face Spoof Detection

S Lokesh Kumar[1], Yamani Sai Asish[1], Sannasi Ganapathy[2]

[1]*School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, INDIA.*
[2]*Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, INDIA.*
slokeshkumar.2018@vitstudent.ac.in, Yamanisai.asish2018@vitstudent.ac.in, sganapathy@vit.ac.in

Abstract **–** Recently, the emerging applications such as banking, mobile payments, face recognition technology are gradually booming and also increases the users count around the world. The extensive deployment of facial recognition systems has drawn close attention to the dependability of facial biometrics in the fight against spoof attacks, in which a picture, video or 3D mask of a real user's face may be used to access facilities or services illegitimately. While a number of anti-spoofing or liveness detection approaches (which identify whether a face is live or spoof when captured) were suggested, the problem is still unresolved because of the difficulty in discovering discriminatory and computer-cost characteristics and techniques for spoof assaults. Existing methods also utilise a full picture or video to determine luminosity. Often though, some facial areas (video frames) are redundant or relate to the confusion of the picture (video). In this paper, we propose a new hybrid deep learning technique called Hybrid Convolutional Neural Network (CNN) based architecture with Long Short-Term Memory (LSTM) units to study the impact in classification. In this technique is applied a non-softmax function for making effective decision on classification. The hybrid approach is implemented followed by a comparative analysis with existing conventional and hybrid techniques used for spoof detection. The proposed model is proved as better than the existing deep learning techniques and other hybrid models in terms of precision, recall, f-measure and accuracy.

Keywords **–** Classification, deep learning, CNN, LSTM, space spoof detection, face recognition.

1. INTRODUCTION

Face identification and verification is difficult task in computer vision and many researchers are working on this area actively. It is difficult research field in the field of computer vision due to the presence of natural and non-intrusive contact that is ideal for the various emerging applications that are doing the identification and verification processes in such applications.

While the development in facial recognition technology has been considerable over the years, ageing of individuals, and a variety of outside illumination provide major difficulties for researchers. While a lot of the work addresses these problems, the absence of any work on face biometric systems is baffling, because face recognition systems are well known to be prone to spoofing attacks. A spoofing attack is performed when a person attempts to fake data and assume the identity of another person. Face biometric technologies are now receiving more attention due to spoofing attempts. The IJCB 2011 face spoofing detection competition, which was held last year, provides proof of this assertion [1].

The biometric features industry is dominated by fingerprint hardware devices that constitute over 92% of the overall market [2]. Experts predict that global facial recognition equipment and licences will grow from $28.5 million in 2015 to $122.8 million by 2024 due to the increase in face identification in mobile systems. The CAGR (or compound annual growth rate) for revenue for face biometrics (covering both visible light facial recognition and infrared-based facial thermography) is 22% [3]. Biometric systems may be accessed and might be the target of various attacks. It is, by far, the assault with the most practical application that is in the realm of possibility. In the same manner as explained before, to enter the system in an illegal manner, a stolen or duplicated biometric characteristic must be applied to the sensor. With this method, it doesn't matter whether someone has security expertise since the attacker just has to fool the biometric characteristic of the authorised user. Therefore, security systems are often designed to defend against spoof attacks using approaches such as hashing, digital signature, or encryption that are mostly useless in such assaults [5]. Reliable anti-spoofing solutions for biometric characteristics, including fingerprints, face, and other biometric features, have recently received considerable attention in the research community [6].

It is possible to fool facial recognition systems by photographing, filming, or 3D modelling the targeted individual and then presenting the resulting image to the camera. Facial pictures are more likely to lead to spoofing attacks since they may be readily found and downloaded. As a general precaution against spoofing, be on the lookout for liveness detection that is designed to identify changes in the body, such as blinks, facial expressions, movements of the lips, and others. Another example is in Pan et al. [7], who used the fact that people blink at an average of 2-4 seconds each blink and came up with an anti-spoofing technique based on that observation. Other measures to reduce recognition include motion analysis, since flat displays and photos (and images, in particular) are believed to be substantially different from actual human faces, which are complex 3D objects [8,9].

Video-based spoofing attacks should be undertaken only if photos are utilised. Skin texture and skin reflectance are also used as anti-spoofing measures. To better understand how facial features differ among images, one could use an intuitive method of investigating the high frequency information in the facial area, since facial displays on mobile phones and smaller pictures likely include less high frequency components compared to actual faces [10,11]. This technique is not likely to succeed when used on photos and movies of better quality, as demonstrated in [12]. Facial texture quality may now be measured using micro-texture analysis, which is also a creative new technique that was only just published [13,14]. The assessments, however, were based on data sets with minimal variance and input images that had good fake face quality. Besides multispectral and multimodal techniques, other defences against face spoofing assaults include facial analysis and multispectral methods. It is much more difficult to defeat a biometric system that uses face recognition in combination with other biometric modalities including gait and voice, than it is to defeat a biometric system that uses a single biometric technology, such as face recognition. Object surfaces in 3D multi-spectral photography may also be utilised to discriminate real faces from imposters [15].

The incidence of spoofing attacks has dramatically increased in the past several years [16]. New privacy and security concerns have emerged in social networking [17]. In order to achieve a better degree of security in social networks, some sort of spoofing detector would be useful. This is very difficult owing to the enormous quantity of data that may be generated each time a social network is used. Many cyberbullying situations arise when someone misuses another person's identity. False user profiles (of the victim) may be established in this instance. A user's profile or personal account on various social networks may be accessed in such a manner that the user's identity is faked by talking to other people or writing comments under the account's name.

The major contributions of this paper are as below:

i. To develop a new architecture with the combination of CNN and LSTM for performing space spoof detection.
ii. The combination of CNN and LSTM with the consideration of non-softmax function is used for performing effective classification.
iii. To achieve highest detection accuracy as well as the evaluation parameters such as precision, recall and f-measure values.

iv. Proved as better than the other hybrid deep learning techniques and normal deep learning algorithms in terms of detection accuracy.

Rest of this paper is organized as below: The relevant works have been discussed by highlighting the contributions, merits and demerits in section 2. Section 3 provides the necessary background details for the proposed model and also explained the working flow of the proposed model. Section 4 demonstrates the performance of the proposed model through experimental results. Section 5 concludes the proposed model with new ideas to enhance further.

2. LITERATURE REVIEW

The rapid development of technological innovation over the last two decades has led to lower-cost access to sophisticated technology gadgets for a significant part of the world's population. Since the early 1990s, facial recognition has been one of the most researched biometric technologies, and for its many benefits compared to other biometrics, it has gained widespread interest [18]. Face recognition is adaptable to nonintrusive collection situations, as well as for use from a distance.

For several computer vision applications, including picture classification and object identification, deep learning has recently shown remarkable improvements [19] [28-32]. While there have been major advances in face recognition, DeepFace, DeepIDs, VGG Face, FaceNet, SphereFace, and ArcFace have also found tremendous success. At least five teams and eight individuals worked together to outperform the face recognition accuracy on extremely difficult face benchmarks, such as LFW or YTF [20]. One of the most popular biometric technologies in the market is face recognition, which is thanks to its quick, accurate, and foolproof identification [21].

Attacks that are spoofed may be identified using several detection techniques. Computationally cheap detection and identification are needed for real-time operation. Conventional pictures are too slow and use unfamiliar recognition techniques for most of the identification methods [22]. Personal face photos of many individuals are available to the public, thanks to social image sharing and social networking services. An imposter could for instance get the photos of real people [23].

Many new techniques to help characterise sceneries, objects, and biometric characteristics have been introduced during the past few years. These characteristics highlight various elements of visual traits, each having a unique use. Many concentrated over the local data while other people focused the complete descriptions. From all the locally available feature descriptors, SIFT, HOG. SURF and LBP are the most frequent feature descriptors utilised for addressing the variability in the picture due to modifies based on the fluctuation in luminosity.

Deep Learning (DL) processes have been shown like in various ML based applications to be an efficient approach to identify spoofing assaults. Many similar studies regard facial distortion to be a binary classification issue, with the system classifying a face as either a genuine user or a false user [24]. For example, CaffeNet and Google Neural Network (CNNs) models are used by authors of [25] and a texture analysis is performed. In order to differentiate a false picture from a genuine imagery applied to CNN for facial liveliness detection, Alotaibi and Mahmood [26] proposed a non-linear diffusion. At the end, an LBP Face Spoofing detection network is suggested in [27] which combines LBP characteristics with profound learning. Although profound learning techniques have enormous potential, they are computationally costly; for training, they need very extensive data sets and the internal complexity of certain applications makes interpreting findings difficult or understanding the algorithm mechanism.

However, given the prevalence of face identification systems, these were the main targets of presentation attacks. Presentation Attacks (PA) is done by malevolent or unintended users, who either seek to impersonate the identity of someone else, or to prevent the system from being recognised (obfuscation attack). However, the exposures of facial authentication systems to PAs were considerably less explored than face recognition performances.

Because biometric applications, such online payment, based on face identification, are prevalent, it is essential, in real life situations, to safeguard true users from impersonation attempts. We will concentrate more on impersonation detection in this survey report. The next section gives a classification of facial PAs. It provides Based on this classification, we will subsequently provide in this article a typology of current face spooking techniques and then a thorough evaluation of these methods, with a thorough comparison of these by taking into account the findings published in the works examined.

3. PROPOSED WORK

The proposed CNN-LSTM model is explained in this section with necessary background information. In this work, the standard classification algorithms such as SVM, k-NN and Random Forest are used for making final decision on instances at fully connected layer of the deep learning algorithm. First, this section explains in detail about the facial spoofing attacks and the proposed methodology is explained with necessary steps.

3.1 Facial Spoofing Attacks

One may take into account that there are essentially two kinds of attacks (PAs). First, with the emergence of the internet and the social media where increasing numbers of individuals upload photographs or videos of their faces, impostors may use such documents to attempt to deceive facial authentication systems for impersonation. Such assaults are often referred to as impersonation attacks (spoofing). The second kind of presentation assault is termed an overwhelming attack, in which a person employs techniques to prevent system recognition (although not necessary via the impersonation of a legal user identity). In summary, while impersonation attacks (spoofing) are usually done by impostors who are prepared to mimic a genuine user, confusion attacks are designed to ensure that the user stays under the face recognition system radar. Despite their completely distinct aims, the ISO standard [16] on biometric PAD lists both kinds of assaults. This article focuses on impersonation assaults, wherein impostors are able either to directly utilise biometric data from a genuine user to launch an attack or build PAIs for face recognition (Presentation Attack Instruments, typically spoofs or fakes

The most frequent assaults on photo attacks and video replay attacks are caused by the growing flow of pictures on the Internet and by low-cost but high-resolution digital equipment. Impostors may easily gather and reuse real user face samples. Photo assaults are carried out by providing a photograph of a real person to the face authentication system. The impostors typically employ many methods. Photo assaults printed include displaying a paper image. In photographic display assaults, on the other hand, the photograph is shown on a digital device screen such as a smartphone, tablet or laptop and delivered to the system. In addition, printed photographs may be chosen to provide some depth to the photo, this strategy is called a warped photo attack. Sliced photo assaults consist of utilising the image as a mask, where the mouth, eyes and/or the nose areas are cut to provide some vivid clues to the face of the imposter behind the photo, such as the mood or eye blinks.

Video replay assaults are more complex in comparison with static picture attacks since they add inherent dynamic information such as blinking the eye, lip movements and changes to facial emotions to imitate brightness. It is possible to differentiate between poor quality 3D masks (for example, created from a printed picture) and high-quality 3D masks (e.g., made out of silicone). The great realism of the "facelike" 3D structure and the realistic replication of a human skin texture in high-end 3D masks makes it harder to spoof 3D masks using conventional PAD approaches (i.e., methods conceived to detect photo or video replay attacks). Today it is still costly to make a high-quality 3D mask and complicated, relying on full 3D capture, which usually requires collaboration with the user. Thus, 3D mask assaults are still far less common than attacks for picture or video playback. However, in the future years 3D mask assaults are anticipated to become more common with the popularisation of 3D acquisition sensors.

PAD techniques are examined for previously unrecorded assaults, most of which have yet to be developed and relied on current methodologies, such as zero/few-shot learning. Obfuscation assaults, which have a very different goal than impersonation attacks (since the idea is to stay unaware of the system), usually depend on facial cosmetics, plastic surgery or facial area occlusion (e.g., using accessories such as scarves or sunglasses). However, in certain instances, obfuscation attacks may also depend on the usage of biometric data from another individual. It varies significantly from common spoofing assaults in its main purpose. In certain instances, however, PAIs may be identical to those employed for impersonational assaults, e.g. another person's face mask.

3.2 Proposed Methodology

The system architecture proposed for the detection of facial spoofs in this paper takes a hybrid approach as opposed to conventional CNN architectures. We take into consideration multiple statistical and deep learning models which are integrated in the CNN architecture to assess the variation in performance. The techniques given below have been implemented which is followed by a detailed description of the proposed architecture that we will be integrating.

A. Convolutional Neural Networks (CNN)

The fundamental component of CNN are evolutionary neuron layers. In image classification tasks, the input to the convolutionary layer is processed using one or more 2D matrices (or channels), and numerous 2D matrices are produced as the output. The number of matrices input

and output may vary. The method for calculating a single output matrix by using the formula given in equation (1).

$$A_j = f(\sum_{i=1}^{N} I_i * K_{i,j} + B_j) \quad (1)$$

First, every I_i input matrix is combined with a kernel matrix $K_{i,j}$. This calculates the total of all matrices and adds B_j to each element in the resultant matrix. Finally, a non-linear f activation function is used to generate one output matrix A_j for each member of the preceding array. Each kernel matrix set is a local function extractor that extracts regional characteristics from the input matrices. The objective of the learning process is to discover sets of kernel matrices K which extract excellent discriminatory functions for categorization of images. In order to train the kernel matrices and biases in common neuron connection weights, the back propagation method that optimises neural network weight may be used here.

In CNN, pooling layer plays a significant role in reducing the functional dimension. A pooling technique should be used to aggregate the neighbours in the convertible output matrices in order to decrease the number of output neurons in the convolutionary layer. The common algorithms for pooling are max and average pooling. The gradient signal should only be redirected to neurons who assist to the pooling output during the error back propagation phase.

Non-linear functions for the activation of the neuron are utilised in ANN. Sigmoid and hyperbolic tangents are nonlinear saturating functions that fall close to zero with the rise in the output gradient. Recent research has shown that non-saturating non-linear functions such as corrected linear $f(x) = max(0, x)$ [ReLU] enhance both training speed and classification efficiency in CNN applications. The ReLU activation function is utilised in the convolutionary layer in our CNN model. Test findings revealed that the activation function ReLU enhances the performance of the classification by 2.5% and the network converges significantly quicker than the sigmoid activation function.

Techniques to speed up and stabilise neural network training, including batch learning, momentum and weight loss, are used. Batch learning is used to enhance the speed and precision of learning. Instead of changing the connection weights after each back propagation, we analyse 128 input samples in a batch and then update the whole batch.

Drop-out method is used to enhance efficiency by altering neurons in each layer during training. A drop-out map with the same neuron size in each layer is randomly initialised to indicate the on or off neuron status at the beginning of each iteration. During training iteration,

the neurons having an off-state are subsequently eliminated from the network by deactivating the activation signal forward propagation and the error signal reverse neuron propagation. For each learning iteration it is comparable to switching between various models such that several models are taught simultaneously. Throughout the test all neurons are enabled, however the activation signal is reduced by average turn-on rate probability during the exercise phase.

B. Long Short-Term Memory (LSTM)

A RNN may make use of the connection between sensor readings, especially in situations when the chronological relationship matters. RNN is capable of capturing sequential information, however it suffers the gradient vanishing issue, which prevents the network from modelling temporal information in a large context frame.

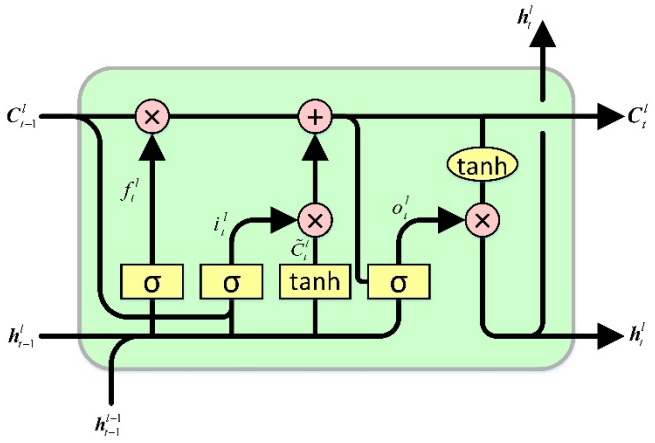

Fig. 2. Long Short-Term Memory Architecture

RNN and LSTM are two varieties of neural network, and LSTM may remove this restriction. Since it uses LSTM cells with particular memory capacity, LSTM has better performance in feature extraction of sequence data than convolutional neural networks. The temporal characteristics in the sequence data are initially extracted using two LSTMs before going through the rest of the pipeline. Each cell in the memory is controlled by an array of input gates, forgetting gates, and output gates. For calculating the activation of each LSTM unit, the following formula that is given in equation (2) is applied:

$$h_t = \sigma\left(w_{i,h}.x_t + w_{h,h}.h_{t-1} + b\right) \quad (2)$$

as the hidden (or input) activation and hidden (or output) activation of a unit are set equal to each other, and a nonlinear activation function is applied, the units' input (or hidden) weight

matrix, $w_{i,h}$, and output (or hidden) weight matrix, $w_{h,h}$, are defined, and the hidden bias vector, b, is then computed. The dimension of the input sample in CNN requires four samples, as opposed to the output dimension of the LSTM layer, which requires three samples. the output of the second LSTM layer is enlarged to match the geometry of the input convolutional layer (samples, 1, time steps, input dimension).

C. Support Vector Machines (SVM)

SVM has a specific characteristic, in that it reduces the loss of empirical identification while concurrently increasing the geometric margin. In order to create a hyperplane that maximally separates the input vector from the feature space, the SVM map input vector to a higher dimensional space. Two hyperplanes are drawn parallel to each other on each side of the original hyperplane to divide the data. This is the hyperplane that maximises the distance between the two parallel hyperplanes. According to this, the generalization error of the classifier will be better the greater the distance between the hyperplanes, regardless of their size. As per the input data, the data points are collected according to the formula given in equation (3).

$$\{(x1, y1), (x2, y2), \ldots . (xn, yn)\} \quad (3)$$

A constant representing the class to which that point belongs, depending on whether or not the value of yn is equal to 1 or -1. n is the number of samples we have. The p-dimensional real vector has an expression of the kind x n. To defend against variance in qualities with more variability, the scaling is critical. By using the dividing (or separating) hyperplane, it visualizes this training data.

$$w. x + b = Output \quad (4)$$

scalar b is given and p-dimensional Vector w is requested. W points perpendicular to the separating hyperplane. Increasing the margin is possible by adding the offset parameter b. The hyperplane must always go via the origin if b is absent, and thus restricts the possible solutions. Maximizing our margin is important, hence we are intrigued in using support vector machines and concurrent hyperplanes. Using equation (5) to describe parallel hyperplanes

$$w. x + b = 1 \; w. x + b = -1 \quad (5)$$

If learning data is linearly separable, we may use these hyperplanes to eliminate any points lying between them, then use the maximisation of the distance between them to optimise

performance. We discover that the distance between the hyperplane and the w-plane is 2 / |w|. Therefore, we aim to decrease |w| to need to make sure that at least one of the data points has been set. Support vectors are values located along the hyperplanes (SVs). A separating subspace with the highest margin defined by $M = 2 / |w|$ that is locations in training data closest to it are specified by the margin defined by the value of M which satisfies the equation (6).

$$yj\ [w\ T\ .\ xj + b] = 1,\ i = 1 \qquad (6)$$

Given that y is greater than $w\ T$, then x must be greater than 1, which means I is 1. (3) This canonical Hyperplane has a maximum margin and is known as Optimal Canonical Hyperplane (OCH). The conditions given in equation (7) must be met by OCH:

$$yi\ [w\ T\ .\ xi + b] = 1,\ i = 1,2\ \ldots\ l \qquad (7)$$

In the space defined by the function φ, training vectors xi are transferred onto a higher-dimensional space. Next, the SVM determines a linear separating hyperplane in this higher-dimensional space with the largest margin. The number greater than zero is the penality parameter of the error phrase. $K(xi\ ,\ xj) \equiv \varphi(xi\ )\ T\varphi(xj)$ is also known as the kernel function. Finding an appropriate kernel function for SVM is also a matter of study. While specific kernels are often used, there are other kernel functions that may be used for generic purposes.

D. K-nearest-neighbor (K-NN)

The algorithm K-nearest-neighbor (KNN) is a learning technique based on an instance. KNN presupposes that every instance has a dot in an n-dimensional space and is describable as an attribute sequence. KNN will pick k closest examples to instance x in the training database in order to classify a new instance x and will utilise k instances to decide instance x.

In the following two areas, KNN has disadvantages. Firstly, since most calculations occur during the classification of a new instance rather than in the training phase, the duration of classification of a new instance is huge and even unacceptable, particularly if the database is large. Secondly, the distance between instances is computed on the basis of all instance characteristics. If there are a lot of irrelevant characteristics, the distance is altered, which has a negative effect on classification accuracy. The courses should thus be properly indexed and based on an active way of learning. We don't have to repeat the prior procedure when categorising a new instance and may use the results of the training phase. Meanwhile, by

filtering irrelevant characteristics, we *attempt* to get a sub-set of all attributes to enhance classification accuracy.

E. Random Forest

Random Forest is an ensemble of learning algorithms based on methods. RF consists of a series of classifiers for tree. Every tree is composed of nodes and edges. The received group classifies new data points through a majority within each classification model 's predictions, as shown in Fig. 2. This approach incorporates a bagging cycle (bootstrap aggregation) and a set of random splits. Each tree is extracted from the data set from a separate bootstrap sample, and each tree categorizes the data. The final outcome is a majority vote between the trees. The random forest algorithm is defined by the following steps:

Step 1: Construct samples of the data from k trees bootstrap.
Step 2: For each of the bootstrap samples grow an unpruned tree.
Step 3: Randomly sample n-try of the predictors at each node, and pick the best split among those factors.
Step 4: Predict new data through a combination of the k tree predictions

3.3 Proposed Architecture

The proposed architecture is shown in fig. 4 that demonstrates the application of SVM, Random Forest and KNN.

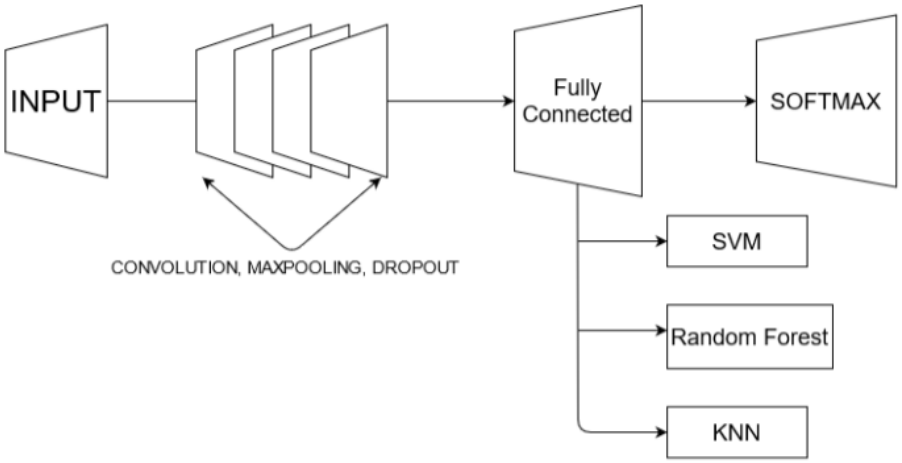

Fig. 4. Hybrid CNN based architecture

Based on the aforementioned algorithms, a hybrid CNN architecture is modified with statistical algorithms along the fully connected layers as shown in Fig. 4. Two separate architectures have

been implemented to analyze the effects on statistical models in the presence and absence of recurrent neural networks. The next section presents a comparative analysis of the overall architecture performance over multiple dataset batches.

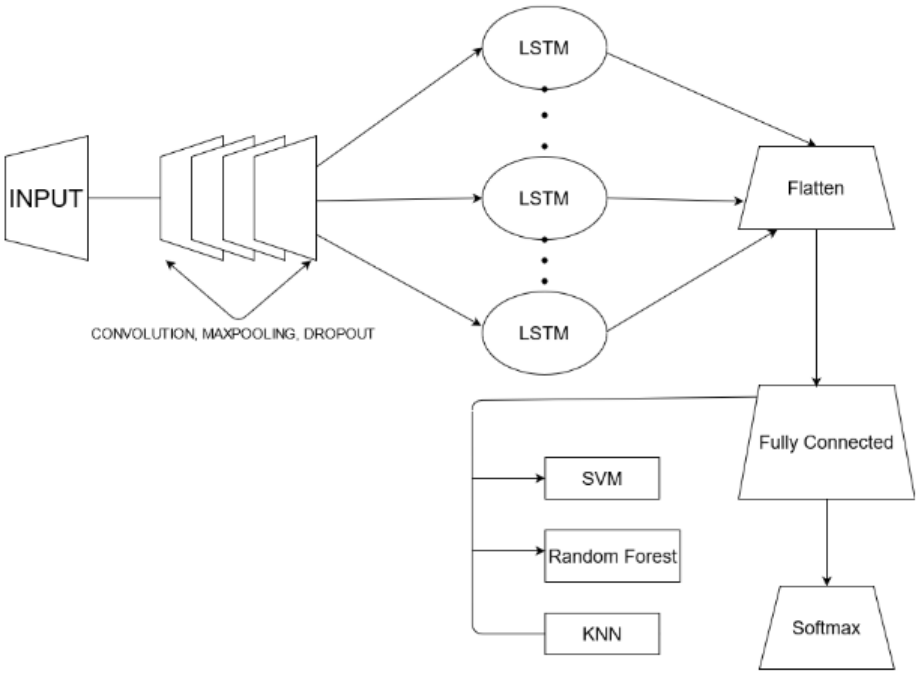

Fig. 5. Hybrid CNN + LSTM based architecture

Fig.5 demonstrates the working flow of the proposed hybrid deep learning technique that combines the CNN and LSTM along with non-sofmax functions. Instead of soft-max function, the proposed architecture applies the LSTM in between the fully connected layer and the convolutional layer. The fully connected layer uses the classifiers SVM, Random Forest and KNN for making effective decision on the input images.

4. RESULTS AND DISCUSSION

The NUAA Photograph Imposter Database [33] has been utilized for evaluating the proposed hybrid model by conducting various experiments. The dataset contains face images output by a face detector taken under different lighting conditions and positions. There are 5105 client images and 7509 imposter images in the dataset which have been divided into five batches for model performance comparisons. Images from each subject are stored in a separate directory as shown in Table 1.

Table 1. Dataset Batch Layout

	Real	**Spoof**
Batch I	420	1224
Batch II	1601	1661
Batch III	1098	1806
Batch IV	920	929
Batch V	1067	1287

Multiple performance metrics have been taken into consideration such as precision, recall, f1-score and support for detailed understanding of how each hybrid model performs with every batch which has variation in class balance. As shown below in Table 2, the metrics have been mentioned for the corresponding algorithms on the batch I of the dataset. On an imbalanced dataset, with a ratio of 1:3 between the classes, the hybrid architectures performed well with hybrid CNN architecture and also hybrid CNN + LSTM architectures providing similar results.

Table 2. Batch I Result

		Precision	**Recall**	**F1-score**	**Accuracy**
CNN Softmax	Real	0.99	0.95	0.97	0.99
	Spoof	0.98	1.00	0.99	
CNN SVM	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN KNN	Real	1.00	0.96	0.98	0.99
	Spoof	0.99	1.00	0.99	
CNN Random Forest	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM Softmax	Real	1.00	0.97	0.99	0.99
	Spoof	0.99	1.00	1.00	
CNN LSTM SVM	Real	0.98	0.97	0.98	0.99
	Spoof	0.99	0.99	0.99	
CNN LSTM KNN	Real	0.99	0.99	0.99	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM Random Forest	Real	0.98	0.99	0.99	0.99
	Spoof	1.00	0.99	1.00	

The second batch has been segregated into a balanced dataset with 1:1 ratio for the real image and spoof image classes. As shown in Table III, the CNN softmax layer provides the least accuracy with 0.97 while the remaining architectures are able to generate full accuracy on the batch.

Table 3. Batch II Result

		Precision	**Recall**	**F1-score**	**Accuracy**
CNN Softmax	Real	0.99	0.95	0.97	0.97
	Spoof	0.95	0.99	0.97	
CNN SVM	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN KNN	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN Random Forest	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM Softmax	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM SVM	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM KNN	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM Random Forest	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	

The third batch follows class ratio as 5:9, generate similar performance attributes as shown in Table 4. The softmax based convolutional neural network is able to provide an accuracy of 0.98 whereas the rest of the hybrid architectures are successful in generating full accuracy for batch III.

Table 4. Batch III Result

		Precision	**Recall**	**F1-score**	**Accuracy**
CNN Softmax	Real	1.00	0.95	0.97	0.98
	Spoof	0.97	1.00	0.98	
CNN SVM	Real	1.00	1.00	1.00	1.00

	Spoof	1.00	1.00	1.00	
CNN KNN	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN Random Forest	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM Softmax	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM SVM	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM KNN	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM Random Forest	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	

For batch IV, the class ratio maintained is 1:1, which leads to full accuracy in the algorithms implemented as the primary study in the paper.

Table 5. Batch IV Result

		Precision	**Recall**	**F1-score**	**Accuracy**
CNN Softmax	Real	1.00	0.99	1.00	1.00
	Spoof	0.99	1.00	1.00	
CNN SVM	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN KNN	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN Random Forest	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM Softmax	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM SVM	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	
CNN LSTM KNN	Real	1.00	1.00	1.00	1.00
	Spoof	1.00	1.00	1.00	

CNN LSTM	Real	1.00	1.00	1.00	1.00
Random Forest	Spoof	1.00	1.00	1.00	

Table 6 shows the performance of various classifiers by considering the batch V of the dataset provides unique performance characteristics as opposed to other dataset batches. Unlike previous batches, hybrid CNN architectures have marginally better as opposed to hybrid CNN + LSTM architecture.

Table 6. Batch V Result

		Precision	**Recall**	**F1-score**	**Accuracy**
CNN Softmax	Real	1.00	0.99	0.99	0.98
	Spoof	0.99	1.00	1.00	
CNN SVM	Real	1.00	0.99	0.99	0.98
	Spoof	0.99	1.00	0.99	
CNN KNN	Real	0.99	0.99	0.99	0.98
	Spoof	0.99	0.98	0.98	
CNN Random	Real	1.00	0.99	0.99	0.99
Forest	Spoof	0.99	1.00	0.99	
CNN-LSTM	Real	0.99	0.98	0.99	0.99
Softmax	Spoof	0.98	0.99	0.99	
CNN-LSTM	Real	1.00	0.98	0.99	0.99
SVM	Spoof	0.98	1.00	0.99	
CNN-LSTM	Real	1.00	0.99	0.99	0.99
KNN	Spoof	0.99	1.00	0.99	
CNN-LSTM	Real	1.00	1.00	1.00	1.00
Random Forest	Spoof	1.00	1.00	1.00	

From table 6, it can be seen that the performance of the proposed architecture-based hybrid model is performed well when applies Random Forest, KNN and SVM. The proposed hybrid model CNN-LSTM with SVM and KNN is achieved 99% accuracy and the proposed hybrid model CNN-LSTM with Random Forest is achieved 100% accuracy for the set of inputs.

5. CONCLUSION AND FUTURE WORK

Given the diversity of materials across spoofing devices, existing anti-spoofing methods are often ineffective. The ability to generalise facial spoofing must be much enhanced before practical applications can be implemented. We concentrate in this article on a wider range of characteristics, the fine-grained movements across video frames, for robust face spotting. The LSTM implemented along with conventional neural network architectures is used to extract temporal elements from the input videos and the LSTM is suggested to increase the movement and attention mechanism to guarantee that LSTM fulfil the dynamic changes shown by people. Furthermore, a comparative analysis is presented amongst some of the most commonly used classification algorithms for spoof detection. Intra-test is conducted on dataset batches to show the efficacy of the proposed approach, and many common methods are used for comparison. The experimental findings show that these dynamic changes in the temporal characteristics are quite useful in improving the capacity for generalizing the method suggested.

DECLARATIONS

Funding: No Funding for this work.

Conflicts of interest/Competing interests: There is no conflicts of interest.

Availability of data and material: Not Applicable

Code availability: -

Authors' contributions: -

Ethics approval: -

Consent to participate: -

Consent for publication: -

REFERENCES

1. Chakka, M.M., Anjos, A., Marcel, S., Tronci, R., Muntoni, D., Fadda, G., Pili, M., Sirena, N., Murgia, G., Ristori, M., Roli, F., Yan, J., Yi, D., Lei, Z., Zhang, Z., Li, S., Schwartz, W.R., Rocha, A., Pedrini, H., Lorenzo-Navarro, J., Castrill´onSantana, M., M¨a¨att¨a, J., Hadid, A., Pietikainen, M., “Competition on counter measures to 2-d facial spoofing

attacks", In: Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA, pp. 1-5, 2011.

2. Girardin, G. Consumers rule: Why the biometrics market is facing major disruption. Biom. Technol. Today 2017, pp. 10–11, 2017.
3. Tractica. Global Biometrics Market Revenue to Reach $15.1 Billion by 2025 | Tractica. Available online: https://www.tractica.com/newsroom/press-releases/global-biometrics-market-revenue-to-reach-15-1- billion-by-2025/
4. Mobile biometrics revenues predicted to boom. Biom. Technol., Vol.10, No.9, pp. 3-12, 2017.
5. Nita, S.L., Mihailescu, M.I., Pau, V.C., "Security and Cryptographic Challenges for Authentication Based on Biometrics Data", Cryptography, Vol.2, No.39, pp. 1-22, 2018.
6. Chugh, T.; Cao, K.; Jain, A.K., "Fingerprint Spoof Buster: Use of Minutiae-Centered Patches", IEEE Trans. Inform. Forensics Security, Vol. 13, pp. 2190–2202, 2018.
7. Pan, G., Wu, Z., Sun, L., "Liveness detection for face recognition", In: Delac, K., Grgic, M., Bartlett, M.S. (eds.) Recent Advances in Face Recognition, Ch. 9. INTECH (2009).
8. Kollreider, K., Fronthaler, H., Bigun, J., "Non-intrusive liveness detection by face images", Image and Vision Computing, Vol. 27, pp. 233–244, 2009.
9. Bao, W., Li, H., Li, N., Jiang, W. "A liveness detection method for face recognition based on optical flow field", In proceedings of 2009 International Conference on Image Analysis and Signal Processing, pp. 233–236, 2009.
10. Li, J., Wang, Y., Tan, T., Jain, A.K., "Live face detection based on the analysis of fourier spectra", In: Biometric Technology for Human Identification, pp. 296–303, 2004.
11. Tan, X., Li, Y., Liu, J., Jiang, L., "Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model", In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, Vol. 6316, pp. 504–517, 2010.
12. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z., "A face anti-spoofing database with diverse attacks", In Proceedings of 5th IAPR International Conference on Biometrics (ICB 2012), New Delhi, India, pp. 34-39, 2012.
13. Bai, J., Ng, T.T., Gao, X., Shi, Y.Q., "Is physics-based liveness detection truly possible with a single image?", In proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), pp. 3425–3428, 2010.
14. Maatta, J., Hadid, A., Pietikainen, M., "Face spoofing detection from single images using micro-texture analysis", In Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA, pp. 56-62, 2011.

15. Zhang, Z., Yi, D., Lei, Z., Li, S.Z, “Face liveness detection by learning multispectral reflectance distributions”, In proceedings of International Conference on Face and Gesture”, pp. 436–441, 2011.
16. Dawson, M., Omar, M., Abramson, J., “Understanding the methods behind cyber terrorism”, In Encyclopedia of Information Science and Technology, 3rd ed.; IGI Global: Hershey PA, USA, pp. 1539–1549, 2015.
17. Liu, F.; Zhu, X., Hu, Y., Ren, L.; Johnson, H., “A cloud theory-based trust computing model in social networks”, Entropy, Vol.19, No.11, pp. 1-21, 2016.
18. Turk, M., Pentland, A., “Eigenfaces for recognition”, J. Cogn. Neurosci., Vol.3, pp. 71–86, 1991.
19. Krizhevsky, A, Sutskever, I., Hinton, G.E., “Imagenet classification with deep convolutional neural networks”, In Proceedings of the NIPS, Lake Tahoe, NV, USA, pp. 1097–1105, 2012.
20. Schroff F., Kalenichenko, D., Philbin, J. Facenet, “A unified embedding for face recognition and clustering”, In Proceedings of the CVPR, Boston, MA, USA, 8–10 June 2015, pp. 815–823, 2015.
21. de Luis-García R., Alberola-López C., Aghzout O., Ruiz-Alzola, J., “Biometric identification systems”, Signal Process, Vol. 83, pp. 2539–2557, 2003.
22. Maatta, J., Hadid, A., Pietikainen, M., “Face spoofing detection from single images using texture and local shape analysis”, IET Biometrics, Vol.1, pp. 3-10, 2012.
23. Chakka, M.M., Anjos, A., Marcel, S., Tronci, R., Muntoni, D., Fadda, G., “Competition on counter measures to 2-D facial spoofing attacks”, In Proceedings of the International Joint Conference on Biometrics (IJCB 2011), Washington, DC, USA, 11–13 December 2011, pp. 1-5, 2011.
24. Yang, J., Lei, Z., Li, S.Z., “Learn Convolutional Neural Network for Face Anti-Spoofing”, Computer Vision and Pattern Recognition, Cornell University, pp. 1-8, 2014.
25. Patel, K., Han, H., Jain, A.K., “Cross-Database Face Anti spoofing with Robust Feature Representation”, In Biometric Recognition; You, Z., Zhou, J., Wang, Y., Sun, Z., Shan, S., Zheng, W., Feng, J., Zhao, Q., Eds.; Springer International Publishing: Cham, Switzerland, Vol. 9967, pp. 611–619, 2016.
26. Alotaibi, A., Mahmood, A., “Deep face liveness detection based on nonlinear diffusion using convolution neural network”, Signal Image Video Process, Vol. 11, pp. 713–720, 2017.

27. Li, L., Feng, X., Xia, Z., Jiang, X., Hadid, A., “Face spoofing detection with local binary pattern network”, Journal of Visual Communication and Image Representation, Vol. 54, 182–192, 2018.
28. T. Chugh and A. K. Jain, "Fingerprint Spoof Detector Generalization," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 42-55, 2021.
29. H. Chen, G. Hu, Z. Lei, Y. Chen, N. M. Robertson and S. Z. Li, "Attention-Based Two-Stream Convolutional Networks for Face Spoofing Detection," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 578-593, 2020.
30. Bowen Zhang, Benedetta Tondi, Mauro Barni, "Adversarial examples for replay attacks against CNN-based face recognition with anti-spoofing capability", Computer Vision and Image Understanding, Vol.197–198, No. 102988, 2020.
31. Neenu Daniel, A.Anitha, "Texture and quality analysis for face spoofing detection", Computers & Electrical Engineering, Vol. 94, No. 107293, pp. 1-21, 2021.
32. Yaowen Xu, Lifang Wu, Meng Jian, Wei-Shi Zheng, Yukun Ma, Zhuming Wang, "Identity-constrained noise modeling with metric learning for face anti-spoofing", Neurocomputing, Vol.434, pp. 149-164, 2021.
33. http://parnec.nuaa.edu.cn/_upload/tpl/02/db/731/template731/pages/xtan/NUAAImposterDB_download.html