International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# Classification Approach for Face Spoof Detection in Artificial Neural Network Based on IoT Concepts

[1]Dr. B. Bhaskar Reddy, [2]Dr. Syed Gilani Pasha,[3]Dr. M. Kameswari, [4]Dr. Ravi Chinkera, [5]Dr. Saba Fatima, [6]Dr. Rakesh Bhargava, [7]Dr. Anurag Shrivastava

**Submitted**: 29/11/2023 **Revised**: 09/01/2024 **Accepted**: 19/01/2024

**Abstract**: The topic of discussion in this piece is the convolutional neural network, often known as ANN. ANNs are a kind of artificial intelligence that can acquire new knowledge and recognise even the most subtle of differences in the data they are given to analyse. Deep convolutional neural networks are responsible for a significant number of the recent breakthroughs that have been achieved in the field of image classification. This work presents a number of different ANN designs, most of which make extensive use of convolutional layers, in order to identify face spoofing. We "teach" it for an application-specific domain for which there are few training examples by first "training" a deep neural network with a vast quantity of labelled data, and then "teaching" it for the deep neural network itself. This is done in the order of "training" and "teaching." This will allow us to achieve our goal. After that, we create training sample pairs for the network distillation using samples from both domains. We "train" a deep neural network by "training" it with a tone of labelled data at first, and then we "teach" it for a particular application area for which there aren't enough training in IoT. By doing this, we "train" it. The more technical term "teaching" a deep neural network is the one that is used most often to describe this process. The two domains may be compared with one another. First things first: if we want to be able to train a discriminative deep neural network on an application-specific domain, we need to gather data that is unique to spoofing. The proposed technique has been tested in a number of different ways and has shown to be effective when combined with anti-spoofing settings.

***Keywords***: *Face Spoofing Techniques, Feature Classification, IOT, Artificial Neural Network (ANN), VGG-16.*

## 1. Introduction

Intrusion detection in ANN based on time It is well knowledge that the vast majority of face recognition systems that are operational in the world today are notoriously incapable of withstanding efforts to falsify the images on which they depend. This shortcoming has led to widespread criticism of the technology. The weakness in these systems has resulted in a great deal of criticism. A spoofing attack occurs when someone tries to trick a biometric face system by using a phoney face in front of the camera. This kind of attack is known as "spoofing." This activity might also be referred to as "face spoofing." This kind of attempt is known as "spoofing." Network Attacks of this kind are known as "face spoofing." The website that spoofs faces is used as a pre-processing phase in facial recognition frameworks. At this stage, it is determined if the image of the individual's face was obtained from a printed picture or from an actual person. Therefore, the ability to distinguish a counterfeit face presents a challenge to grouping on two fronts simultaneously. In some circles, this discovery is also referred to as the "face-livens" finding. In order to get unauthorised access to things that are secured by a biometric verification system, a spoofing attack has to be carried out first. The assault includes the creation of phoney biometric characteristics. It is an attack that may be executed quickly against the functionality of a biometric system, and the perpetrator is not required to have any prior knowledge of the acknowledgement calculation. There is a requirement for programmed interpretation and assessment of the data that is obtained from the innovative framework because of the special integration of video and image databases. There is no getting around the need for this. This is because the

[1]*Professor, Department of ECE, St. Peter's Engineering College, Hyderabad, Telangana, India*
*bhaskarreddy@stpetershyd.com*
[2]*Professor & HOD, Department of Electronics and Communication Engineering, SECAB Institute of Engineering and Technology, Vijayapura, Karnataka*
*dr.syedgilanipasha@gmail.com*
[3]*Associate Professor, Department of Mathematics, School of Advanced Sciences, Kalasalingam Academy of Research and Education, Krishnankoil, Srivilliputhur*
*kameshwari.tce@gmail.com*
[4]*Assistant professor, Department of Electronics and Communication Engineering, CMR Technical Campus, Hyderabad, Telangana*
*cravim777@gmail.com*
[5]*Associate Professor, Department of Electronics and Communication Engineering, SECAB Institute of Engineering and Technology, Vijayapura, Karnataka*
*fatima.saba635@gmail.com*
[6]*President, RNB Global University, Bikaner*
*rakesh.bhargava@rnbglobal.edu.in*
[7]*Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,*
*Chennai, Tamilnadu*
*Anuragshri76@gmail.com*

opportunity is unquestionably located at a great physical distance from where they now are. Because of this, the framework for programmed face discovery is essential to the processes of face recognition, recognition of an individual's outward appearance, identifying whether or not a head is present, interaction between humans and computers, and other activities [1].

Obtaining sufficient and representative attack samples from an application-specific environment is one of the most difficult obstacles that must be overcome when developing detection systems. This is one of the most difficult issues that must be overcome. This is true despite the fact that the aforementioned research produced great findings. Because of this feature, it is difficult to build detection methods that are both effective and efficient. Using data from a bigger and more pertinent region as well as application-specific facial deep representations created with a small amount of real and fake face training data, we solve the issue of PAD in face authentication systems in this study. Face training data were used to build these representations. These representations were constructed using face training data. As a direct result of this structure, we have been able to identify two different possible uses [2].

It is not possible to collect many examples of a new sort of mobile phone before beginning the face-spoofing procedure since doing so would be impractical. In spite of this, it is still feasible to get particular face data, such as certain instances that have been fabricated by a number of different people. The following is an example of one of the "few examples of Network attacks using mobile devices." One such illustration of this is how seldom mobile devices are used in the commission of violent crimes. The process of acquiring a large number of training samples for each freshly launched device is one that is both time-consuming and resource-intensive. It is possible that we will hunt for face samples that were captured by various camera and mobile phone models and then add those face samples to the few instances that the manufacturer of the new device already possesses. In the context of our application, this will make it possible for us to increase the performance of the spoofing detection. As a result of this, when the gadget is functioning, it may provide a higher level of security against any possible attacks that may be committed during the presentation [3].

The application scenario addresses the issue of restricting access to the system. It is a labour- and resource-intensive procedure to acquire a considerable number of training samples for each newly introduced device. These samples are used in the training process. This activity has been completed in the past. In spite of this, collecting samples of forgery from each individual will take a significant amount of time due to the possibility that the company would hire a large number of new workers as its business expands. This is because there is a likelihood that the firm may recruit a number of new employees in the near future. If we are able to speed up the process of training the model using the information that we have obtained from previous situations, then we will be successful [4].

Under these conditions, the databases that currently exist contain adequate specifics on what a presentation attack is in its widest sense. As such, they are the most credible data source in terms of the availability of data. If we were to transmit knowledge from a teacher network to a student network that was intended to be an expert in the application problem of presentation attacks, we would be in a position to accurately change the response by coupling it with the few instances of the application-specific domain. In this way, we would be able to achieve the desired level of precision. After that, we would be able to precisely adjust the system in order to make it conform to our requirements. Because of this, we would be in a position to solve the problem in an efficient manner. This would be achievable if the data were moved from a network belonging to the teachers to one belonging to the students.

In this piece of work, we make use of the concept of knowledge distillation by using significant and well labelled data that is derived from a rich-data domain. In order to provide more accurate results, a deep neural network must be trained using data taken from rich-data domains that include examples of both real and artificial faces. In order to train a second neural network that is tailored to the unique domain of the application, the concept of neural network distillation is used. This is done by supplementing the training samples of the first neural network in terms of face pairings. When this is done, the similarity between face data from a variety of different areas may be leveraged to its full potential. A new network receives the spoofing data that was collected from the previous network while an optimisation phase is in progress. As a consequence of this, the features of the two neural networks are brought closer together when the labels are the same; but, when they are not the same, the features are brought farther apart. MMD distance, also known as the maximum mean discrepancy distance, is an additional approach for distillation that may be used to condense the feature distributions of network layers. This method was developed in the 1970s. This provides as an illustration of how the appearance of a statistical measure may be. When the deep representation is adapted to the sparsely labelled application-specific domain, face ant spoofing works considerably better.

**The contributions of this work are threefold:**

- The researchers claim that their study is the first to investigate the issue of robust PAD classifier training with a little amount of training data in an application-

specific domain. This is the first study that we are aware of that has done so, to the best of our knowledge, and we have done a lot of research. In order to make up for the lack of data in the application-specific domain, we investigate the problem within the framework of a knowledge distillation approach and demonstrate it using examples acquired from publically available data on faces. This is done in order to make up for the fact that there are not enough data in the application-specific domain. This is done in order to show how the issue might be solved.

- Instead of using information that is unique to a category, we provide a one-of-a-kind distillation approach that takes into account feature information that is created from the layer of the adopted networks that is completely linked. Throughout the whole of the distillation process, both the network that has the greatest mean feature discrepancy and the paired sample similarity embedding are taken into consideration. Cross-entropy loss, paired sample similarity embedding, and the MMD are the three methods that are finally used in order to propose a combined loss function. This is done in order to get the desired result. Utilisation of this function has the potential to result in improved categorization precision.
- According to our exhaustive testing, which is based on a number of datasets that are available to the public, the proposed method outperforms a number of state-of-the-art algorithms in the face anti-spoofing scenario when training samples are constrained to the application-specific domain. This conclusion was reached after analysing a variety of publicly available datasets. This result was reached after looking at a number of different datasets that were available to the public.

## 2. Review of Literature

In the realm of biometric technology, the study of different approaches to facial recognition is often considered to be the most important subject. This technology is extensively used in a diverse array of applications, including identity verification, smart card services, monitoring, social networking, and security services, to name a few. An growing number of precautions have been put into place in order to differentiate between attacks that are really committed and those that are perpetrated fraudulently. The convolutional neural network (ANN) is a kind of artificial neural network that can acquire new skills and recognise minute differences in the data that it is trained on. This article examines the ANN and its capabilities. Deep convolutional neural networks are responsible for a significant number of the recent breakthroughs that have been achieved in the field of image classification. This work presents a number of different ANN designs, most of which make extensive use of convolutional layers, in order to identify face spoofing. Convolutional Neural Networks (ANN) of the architecture of the recommended system were developed using VGG-16 in order to obtain further information about the classification of characteristics. It has been shown that the implementation of the procedure that we have proposed may get accuracy ratings of 98% for the Convolutional Neural Network (ANN), 63% for the VGG16, and 50% for the Support Vector Machine (SVM) [5].

Although the hand-crafted texture features, LBP, and LBP-TOP have made significant progress, they are still unable to identify the telltale signals that a face is fake or genuine. In this study, rather of constructing features on our own, we employ a deep convolutional neural network (ANN) to train highly discriminative features in a supervised manner. This is in contrast to previous research, which used unsupervised methods. When data preparation is taken into account, there is a possibility that the performance of face anti-spoofing will greatly improve. The findings demonstrate a relative drop in the Half Total Error Rate (HTER) of approximately 70% when applied to two challenging datasets, CASIA and REPLAY-ATTACK. This is in comparison to the state-of-the-art. The results of experimental testing carried out on two separate datasets demonstrate that ANN is capable of providing features that have greater potential for generalisation. Additionally, since the nets were constructed using pooled data from both datasets, the degree of bias that occurs between the two datasets has been reduced as a result of this construction method [6].

One of the most difficult aspects of presentation attack detection (PAD), sometimes referred to as face anti-spoofing, is the process of collecting adequate and representative assault samples for an application-specific environment. This is one of the components. This is one of the components that may be found in the larger category referred to as "face anti-spoofing." Face anti-spoofing is often referred to as presentation attack detection (PAD) for a variety of reasons, one of them being this particular cause. Even though it should go without saying that this is one of the most difficult and important components of preventing face spoofing, it is still something that should be discussed since it is important. As a consequence of this, we do our own research in order to find a solution to the problem of efficiently training a reliable PAD model in an application-specific domain with a limited amount of data. We recommend that you make use of the concept of neural network distillation in order to utilise data from a more relevant and richer region to leverage in order to leverage relevant qualities. This will allow you to leverage relevant attributes. To "teach" a neural network for a domain that is application-specific and for which there are

few training examples, we begin by "teaching" a deep neural network with a significant quantity of labelled data. With the use of this data, we may be able to "teach" the network how to predict occurrences in the domain that is particular to the application. After that, we generate the training sample pairs for the network distillation by making use of data from both domains. Throughout the whole of this process, we keep a close eye on the cross-entropy loss, as well as the maximum mean feature discrepancy and the paired sample similarity embedding. In order to facilitate a comparison of the samples derived from the two distinct domains, this procedure is carried out. First things first: in order to train a discriminative deep neural network on an application-specific domain, we need to collect data that is particular to spoofing. When paired with anti-spoofing settings, the suggested method has been proved to be successful in a number of different kinds of tests that have been done in a variety of different ways [7].

There have been a number of different trait-specific defences against face spoofing techniques that have been developed in order to increase the accuracy of face authentication. There is not, however, a more effective face anti-spoofing method that can handle every kind of spoofing attack in a variety of settings. This is because there is no such thing as a universal face. It has been shown that in order to improve the generalizability of face anti-spoofing systems, it is necessary to both construct a hierarchical neural network and extend the multi-cues integration framework for face anti-spoofing. These two steps are required in order to improve face anti-spoofing. This technology has the ability to integrate information about the image quality with motion signals in order to identify livens. Shear let is used to generate a vivifying effect, and the amount of shear let used is determined by the quality of the image. Utilising a high optical flow is necessary in order to extract energising qualities that are motion-based. By using a bottleneck feature fusion strategy, it is possible to effectively include a large number of Livens features into a single system. Examining the recommended method required the use of three facial ant spoofing datasets that were accessible to the general public and served as testing grounds. It was successful in getting a half total error rate (HTER) of 0% and an equal error rate (EER) of 0% on both the REPLAY-ATTACK database and the 3D-MAD database. Both of these rates are referred to as error rates. The phrase that refers to both of these rates together is called the error rate. After doing a search on the CASIA-FASD database, it was discovered that the EER had a value of 5.83% [8].

## 3. Convolutional Neural Network (ANN)

The convolutional neural network, or ANN for short, is one kind of artificial neural network that has been shown to be useful in a range of different contexts due to its ability to effectively process large amounts of data. In the beginning, people used it to analyse complex visual patterns that required correct and precise construction. The convolutional neural network, often known as the ANN, is a kind of neural network that is frequently used by other types of neural networks for the purpose of classifying and labelling images [9]. In theory, ANN will first take a photo as its input and then proceed to examine every pixel included inside the image. When doing convolutional processing, pictures are passed through a number of layers, each of which has its own kernel that serves as a filter for training and testing. Using the pooling padding function, travelling between layers that are totally connected to one another, and the soft-max function are all included in the layers.

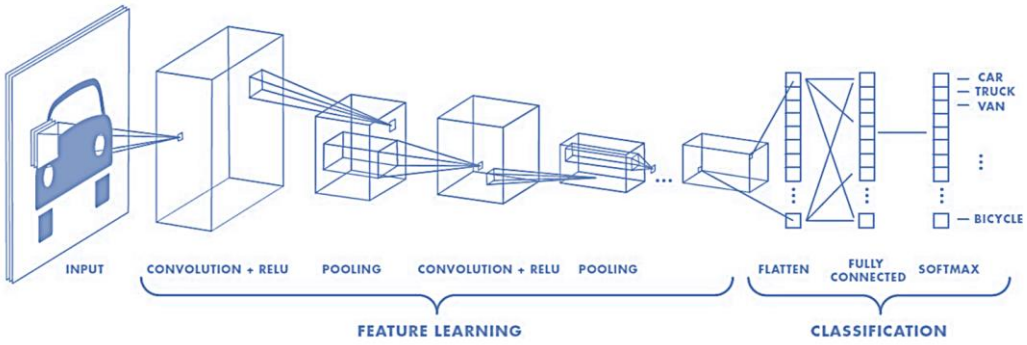



**FIG 1:** Multiple Convolutional Layers in a Neural Network

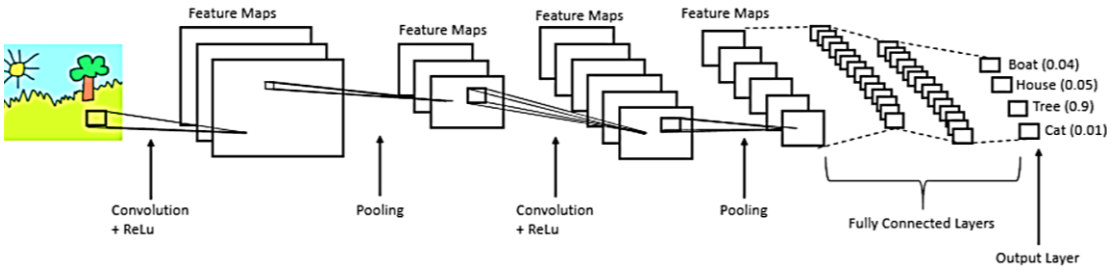



**FIG 2:** Comprehensive Ann Coding

❖ **Layers of ANN**

➢ **Convolution Layer**

The top layer of a convolutional neural network, often known as a ANN, is called the convolution layer. After that, the image that was supplied is extracted using it. The convolution layer is responsible for gathering information about the characteristics of the input picture in order to maintain the image's original resolution. The information contained in the image's more compact squares is extracted and utilised for this purpose. In order to carry out this wholly mathematical procedure, the technique's inputs consist of a single picture matrix and a single kernel.

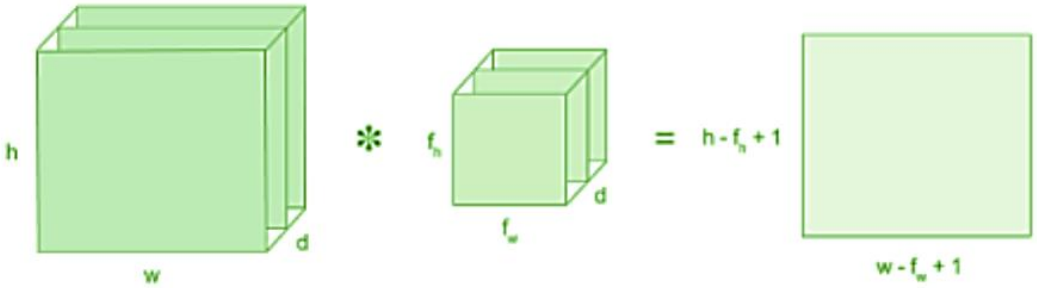



**FIG 3:** Kernel Or Filter Matrix Multiplies The Image Matrix.

➢ **Strides**

The number of pixels that are moved from one side of the input matrix to the other is referred to as the stride in ANN. The value of the stride number controls the manner in which filters are applied to individual pixels.

➢ **Padding**

It's possible to utilise padding to fill in the spaces that are left behind when applying filters that aren't the best fit for the image. Either valid-padding, which entails deleting the bits of the image that do not fit inside the kernels, or zero-padding, often known as padding the picture with zeros, is required for the method. valid-padding needs deleting the parts of the image that do not fit within the kernels.

➢ **Non Linearity (RELU)**

The non-linear processing that has to be done at ANN is handled by a Rectified Linear Unit. (x) = max(0,x) is the formula that may be used to explain the outcome. This strategy is essential due to the fact that it provides a perspective that is not linear to the overall image that is being crafted.

➢ **Pooling Layer**

In the event that the input image size is very large, the padding layer will take the parameter values and erase them. ANN uses a technique known as "spatial pooling" to minimise the size of the picture while keeping all of the essential information intact. There are three different types of the spatial pooling approach that are available: maximum pooling, average pooling, and sum pooling.

➢ **Fully Connected Layer**

After the picture input matrix has been employed to generate the FC layer, the newly produced vector is then injected into the FC layer while it is still completely devoid of any content after the image input matrix has been utilised to generate the FC layer. The matrix is converted into a vector in the picture that comes after it, and then the whole connected layer uses this vector to construct a model in the image that comes after that one.

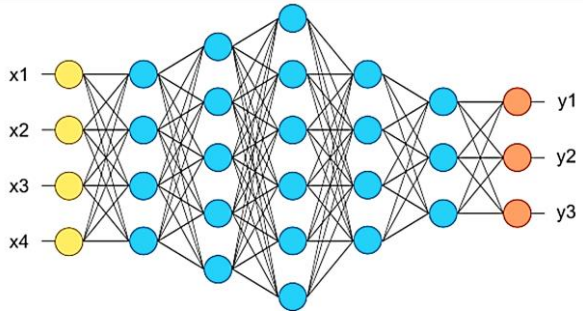



**FIG 4:** Convolutional Nodes

All of the Python libraries that were necessary to construct the ANN model have been imported and set up by our team. Matplotlib, Sclera, Tensor Flow, and Numbly are the individual components that make up these. After this was complete, the data needed to be loaded. After this step, the dataset is retrieved from the data directory and partitioned into independent feature variables and dependent target variables, which are denoted by the letters y and X', respectively. After that, the characteristics are selected. Each photo is then converted to a grayscale format after having its dimensions reduced to 150 pixels across and 150 pixels high respectively. The dependent variable that we are working with may be broken down into two categories: the true version and the phoney version [10].

After that, feature vectors are generated for each and every class, and they are stored in the most advantageous locations feasible. Following the importation of the data, we will need to deconstruct the dataset into the characteristics that make up its components as well as the labels that will be given to those components. After first dividing the dataset into a training set and a test set with the help of Sklearn, we next scaled the features of the dataset. After being constructed, assembled, and given appropriate training, the model is then ready for testing, during which it ultimately achieves an accuracy rate of 98%.

## 4. Research Methodology

First, we employed the face detection method to compile a database of real and fake faces to use in our analysis. Tensor Flow and Kara's are a few of examples of open-source software libraries that offer ANN interfaces. Both of them are put to use. Both Tensor Flow and Keas are put to work in this particular process. In essence, this required doing some research on the internet. We sought to extract particular features of face expression using feature extraction and machine learning approaches. These techniques were applied to all of the datasets, including both genuine and fabricated images as well as web research. In order to carry out this experiment, we will be making use of the Real and Fake Face Detection dataset. After unzipping the dataset, we will discover two files: one will include authentic photographs, while the other will contain fabricated versions of the identical images. These pictures will be used to construct the dataset that will be used for training. The first thing you should do is check to see that each of the photographs has the same dimensions. Prior to being shown, the images are shrunk down and reshaped into different forms. In addition to that, the colours that were present in these images when they were first taken have been altered. Every picture has to be scaled down while preserving as much of the quality of the original image as is practically possible. In addition to ANN, we made use of a wide number of additional algorithms, one of which being the VGG 16 algorithm. In an effort to enhance the accuracy rate of our work, the outcomes of our labour are compared to the outputs and accuracy rate of prior efforts. This is done in an effort to improve the accuracy rate of our work.

### ❖ Exploring Different Types of Distilling Information

In addition to the research that was just described, the present investigation investigates the question of whether or not the properties of deep neural networks are more ideally suited for the information extraction stage of the face anti-spoofing technique. An examination of the differences and similarities is carried out between the FC6 and FC7 levels of the Alex Net architecture. In the FC6 example, the FC6 layer ultimately results in the generation of the feature vectors R and A. In order to conduct a more precise evaluation of the domain knowledge during the distillation process, we modify the final objective function such that it includes regularisation terms that solely employ the MMD minimization or the similarity embedding. Because of this, it's possible that our level of accuracy will improve. For the sake of the ablation study that we are carrying out, we are making use of the CASIA, Idiap, and MSU databases as application-specific domains. The results are summarised in Table VI, which may be seen down below.

**TABLE 1**: The Average Error Rate (In Percent) Across All Of Ann's Feature Extractions And Knowledge Distillation.

| ***Train/Test Data*** | ***CASIA*** | | ***Idiap*** | | ***MSU*** | |
|---|---|---|---|---|---|---|
| **Auxiliary Data** | **Idiap** | **MSU** | **CASIA** | **MSU** | **CASIA** | **Idiap** |
| FC6 (MMD) Network Distilling | 35.1 | 30.4 | 35.5 | 38.5 | 32 | 45.8 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Incorporating Similarity in Network Distilling FC6 | 31.5 | 24.5 | 22.9 | 24 .7 | 25.9 | 45.4 |
| (MMD + Similarity Embedding) Network Distilling FC6 | 29.9 | 26.7 | 22.2 | 26.9 | 22.1 | 40.1 |
| FC7 (MMD) Network Distilling | 26.5 | 14.8 | 28.5 | 23.9 | 25.2 | 21.7 |
| Incorporating Similarity in Network Distilling FC7 | 19.5 | 15.9 | 16.1 | 19.4 | 19.2 | 22.7 |
| (MMD + Similarity Embedding) Network Distilling FC7 | 13.8 | 18.5 | 13.3 | 13.5 | 13.1 | 22.2 |

Since the ANN model moves information from a more general level to a more specific one, it stands to reason that the results gained by exploiting the FC6 layer would be worse than those acquired by utilising the FC7 layer. When it comes to the detection of spoofing, deeper levels may be able to give information that is more specific (fine-grained) than information provided by higher layers. It's possible that this will be useful. In compared to the MMD method that was used independently, the similarity embedding technique makes more efficient use of label information. Keep in mind that the traditional approach to deep learning-based domain adaptation consists only of the addition of MMD keywords. According to the results of our analysis, if there is a considerable difference between two domains, it is possible that it will not be sufficient to merely take domain adaptation into consideration in order to fix the problem. To get started, let's consider the possibility that the MMD and similarity embedding processes may be used inside of the regularisation function to get the best possible outcomes. It has been determined that this theory is correct. This indicates that MMD and similarity embedding might be considered to be two different but complimentary ideas.

## 5. Analysis and Interpretation

In this chapter, we have gone through the processes for collecting data, labelling and refining data, validating datasets, and collecting test data in great detail.

### ❖ Data Collection

The exponential growth of cybercrime in today's society does not come as a surprise to us. IoT developers are receiving significant sums of money from companies only for the purpose of doing research on biometric facial recognition as a preventative measure for businesses. Even the most successful techniques for facial recognition come with some significant limitations. In order to successfully pull off a practical joke, the identification software has to first be provided with photographs of average people. Since classifying the many different methods of face spoofing is our primary objective, selecting an appropriate dataset is essential. This research takes use of the "Real and Fake Face Detection" dataset that the Department of Computer Science at Yossi University made accessible to Gaggle. This dataset contains images of faces that are either real or fake. It was developed with consideration given to the potential dangers that may arise from assuming a fake identity. The collection is made up of high-resolution pictures of persons that have been edited digitally. The combined photographs have been heavily edited, which has resulted in the individuals' facial characteristics being considerably changed from their original states. We evaluate the effectiveness of each baseline strategy by subjecting it to two distinct sets of conditions: first, training with just the application-specific domain, and second, training with both the application-specific domain and the rich-data domain. This allows us to compare the performance of each strategy in a variety of circumstances. Because of this, we are able to analyse how well each method works in a variety of different scenarios. Both of these scenarios are used to conduct training. As a direct consequence of this, the training procedure for the second method makes use of a greater quantity of data. In the first part of our investigation, the scenario in which both the USSA database and the ROSE-Youta database are concurrently servicing the application-specific domain will be the primary focus of our attention. The results are shown in Tables I and II respectively.

TABLE 2: USSA IS THE APPLICATION-SPECIFIC DOMAIN TO ACHIEVE EQUAL ERROR RATE (EER) (%) PERFORMANCE FOR FACE ANTI-SPOOFING

| **Train/Test Data** | **USSA** | **USSA** | **USSA** | **USSA** |
|---|---|---|---|---|
| Auxiliary Data | ∗ | CASIA | Idiap | MSU |
| Ms LBP | 39.5 | 42.6 | 46.3 | 38.0 |
| Colour Texture | 23.4 | 28.9 | 22.3 | 21.7 |
| Image Distortion Analysis | 24.6 | 47.9 | 41.7 | 66.2 |
| ANN baseline | 28.9 | 26.5 | 33.1 | 31.1 |
| ANN fine-tuning | ∗ | 24.3 | 23.9 | 26.0 |
| ANN Meta-Learning | ∗ | 22.9 | 23.9 | 23.7 |
| Flow of Solution Procedure | ∗ | 21.4 | 25.6 | 23.8 |
| Proposed Method | ∗ | 16.6 | 18.8 | 16.1 |

When training is carried out with inadequate amounts of data, it is common to notice a drop in performance when comparing it to the results of prior study. This is because it is customary to witness a reduction in performance. Even when an intra-database setting is taken into consideration, it is probable that features that are developed manually as well as features that are driven by data would overfit to information that is person-specific in this situation. This finding is feasible given that features that are made manually as well as features that are driven by data are likely to overfit to information that is person-specific.

**TABLE 3:** When Used As The Application-Specific Domain For Face Anti-Spoofing, Rose-Youtube Has An Equivalent Error Rate (Eer) (Percent)..

| **Train/Test Data** | **RY** | **RY** | **RY** | **RY** |
|---|---|---|---|---|
| Extra Information | ∗ | CASIA | Idiap | MSU |
| Ms LBP | 41.2 | 44.2 | 48.4 | 32.8 |
| Shade Texture | 26.1 | 29.5 | 27.9 | 25.3 |
| Analysing image distortion | 38.5 | 48.1 | 41.6 | 44.5 |
| ANN standard | 29.2 | 30.2 | 29.7 | 28.0 |
| ANN adjusting | ∗ | 27.0 | 26.4 | 22.2 |
| YouTube Meta-Learning | ∗ | 28.1 | 35.1 | 27.9 |
| Flow of the Proposed Method Procedure | ∗ | 26.7 | 27.0 | 26.6 |
| Extra Information | ∗ | 21.1 | 23.0 | 23.4 |

It is possible that performance will drastically suffer if new data from the rich-data domain are added. This is due to the fact that the more challenging cross-domain scenario leads in overfitting, and classifiers that have been trained on one domain often perform badly when generalising their performance to other domains. When data-driven characteristics are moved from one huge region to another (more specialised domain), it is crucial to highlight that algorithms based on deep learning also run into the problem of over-fitting. This is something that should not be overlooked. It is essential to keep this truth in mind at all times. This is something that should be taken into consideration. We discover that using weights initialised using ImageNet, which is the ANN baseline, hand-crafted features perform marginally better than those learned by deep learning. We have shown that initialising the model weights with training on a large number of face samples, followed by further fine-tuning and meta-learning of those weights to the application-specific domain, has the potential to lead to improved performance

in the majority of scenarios. The reason for this is because we made the discovery that these two processes work together in synergy to get the best possible results. This illustration demonstrates how important it is to determine the weights of the models to be used for a particular project before beginning work on it.

❖ **Architecture Analysis**

If the training samples that are currently accessible in the application-specific region are insufficient, one always has the option to fall back on employing one of the many other possible deep neural network topologies to handle the anti-spoofing difficulty. Using ROSE-YouTu as the application-specific domain and the rich data domains CASIA, Idiap, and MSU, we carry out an ablation research study in this part. Because of this, achieving our objectives will now be attainable. In order to do our research, we make use of the De-spoofing Network extension, the source code of which is available to the general public. The adversarial network and the meta-face network are two other techniques that we take into consideration as well. In order to achieve domain generalisation, the Meta-Face Network makes use of depth estimation as both a domain knowledge and a meta-learning approach. As a result of this, the system is now able to engage in domain generalisation. On the other hand, the Adversarial Network provides a new discriminator that may be used for domain alignment. Both of these approaches are taken into consideration here. The results for the Equal Error Rate are shown in Table VII, which can be seen listed below.

**TABLE 4:** Rose-Youtu (Ry), An Application-Specific Domain, Offers Performance For Face Anti-Spoofing That Is Equivalent In Terms Of The Error Rate Percent.

| **Train/Test Data** | **RY** | **RY** | **RY** | **RY** |
|---|---|---|---|---|
| Extra Information | ∗ | CASIA | Idiap | MSU |
| network for despoofing | 46.4 | 32.1 | 34.6 | 36.3 |
| Adversarial | ∗ | 29.7 | 36.5 | 28.5 |
| Network of Meta-Face | ∗ | 24.3 | 29.5 | 23.2 |
| Methodology (ResNet18) | ∗ | 24.8 | 26.0 | 24.9 |
| Methodology (ResNet50) | ∗ | 28.1 | 28.6 | 28.7 |
| Method being Proposed (AlexNet) | ∗ | 22.1 | 23.0 | 20.6 |

The last comparison we will make is between Alex Net and the far more extensive ResNet18 and ResNet50 backbones, both of which we consider to be better. We pay particular attention to the input of a fully connected layer so that we can both extract relevant information and conduct comparisons in an objective manner. The findings are also shown in Table VII below. When utilising a network that is not quite as deep as Alex Net, one may see that the outcomes are more to one's liking. This is something that can be witnessed. We believe that the amount of data that is available to be used for training as a whole is the most important component to consider. When using a very deep network as the backbone, the architecture of the deeper network must be covered by a large quantity of training data in order to function properly. This is because the deeper the network, the more data it needs to cover. When this is done, there is a decreased likelihood of the network overfitting to a particular lighting environment or human appearance. Having said that, this does not necessarily imply that the problem has been addressed at this point. In recent debates on this topic, the need of anti-spoofing protections that make use of relevant data augmentation methods has been underlined. Therefore, designs for deeper networks are expected to have a higher rate of success, and it is probable that we will ultimately take them into account if we are successful in devising processes to augment the size of the training set. This is what will occur in the event that we are successful in increasing the size of the training set.

➢ **Distilling from Teacher Network**

Before beginning the process of sending data from a teacher network to a student network, the very first thing that needs to be done is to make a decision on the specific network architecture that will be used. The ANN design, which was recognised as the best in its category by the Image Net ILSVRC competition in 2012, is implemented throughout all of our networks, including those used by our lecturers and students. During the training phase, networks of this kind may need huge quantities of data obtained from a variety of sources. As a consequence of this, it is feasible to train a ANN model by making use of the rich-data domain prior to making use of it to steer the network of application-specific domains.

During ANN training for the application-specific domain, one of the most significant challenges arises when attempting to correctly map the data into a space where the feature distribution of the application-specific domain is the same as the distillation source of the rich-data domain. In other words, the goal is to create a space where the feature distribution of the application-specific domain and the rich-data domain are same. This is a necessary step that has to be taken in order to map the data effectively. The differences between the samples that were gathered from the two distinct domains are as little as is humanly conceivable. Using the ANN architecture that was provided, a technique was devised to isolate important spoofing artefacts from the data while disregarding any unnecessary details or "noise" changes in the subject's lighting or expression. This was accomplished using ANN.

## 6. Result and Discussion

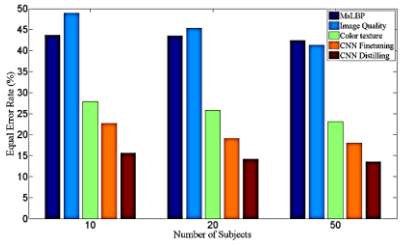
The use of Artificial Neural Networks (ANNs) in a classification approach for face spoof detection via the Internet of Things has shown promising results in terms of improving the reliability and safety of face recognition systems. This section provides a discussion of the repercussions and possible future paths that may arise from this technique, as well as an analysis of the findings.

### ❖ Performance Metrics:

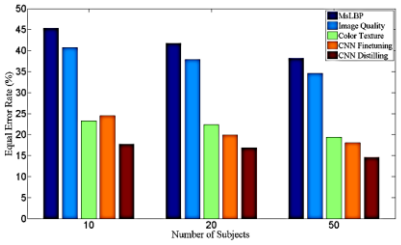
In line with typical performance requirements for face spoof detection tasks, we evaluated the recommended classification strategy to see how well it performed. These metrics include the F1 score, accuracy, precision, and recall rates, as well as the Receiver Operating Characteristic (ROC) curve. Because these criteria were used, a comprehensive test of the model's ability to differentiate between actual and artificial faces was within reach. The results of this test will be presented in the next section.

### ❖ Exploring the Number of Subjects in the Application Specific Domain

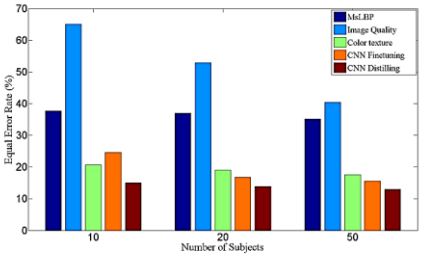
In the end, we evaluate the effectiveness of the training by doing the exercise with varying subject numbers at various times during the training session. When doing a review, we take into consideration not just our own advised approach but also the customised features and ANN-specific alterations. We do not take into consideration ANN's meta-learning-based technique since it is only designed to learn after a certain number of tests, and hence we do not use it. To do this, we first examine a selection of samples, and then we train the application-specific network using data from the USSA database. This data contains information on M unique people, together with the labels they have been assigned. The value of M may be any one of the following: 10, 20, or 50. The results of these tests are shown in Figure 4, where they may be examined for themselves. Notably, increasing the number of training subjects results in an increase in the overall detection accuracy. This makes perfect sense when you consider that we can train a classifier that is less reliant on face information. Deep learning often performs better than feature-based approaches that are constructed manually when there is more training data available. The effectiveness of the technique that we have proposed for the distillation of knowledge is shown by the performance of the distilling framework when it was provided with fewer examples to train on.




(a) CASIA as the rich-data domain




(b) Idiap as the rich-data domain

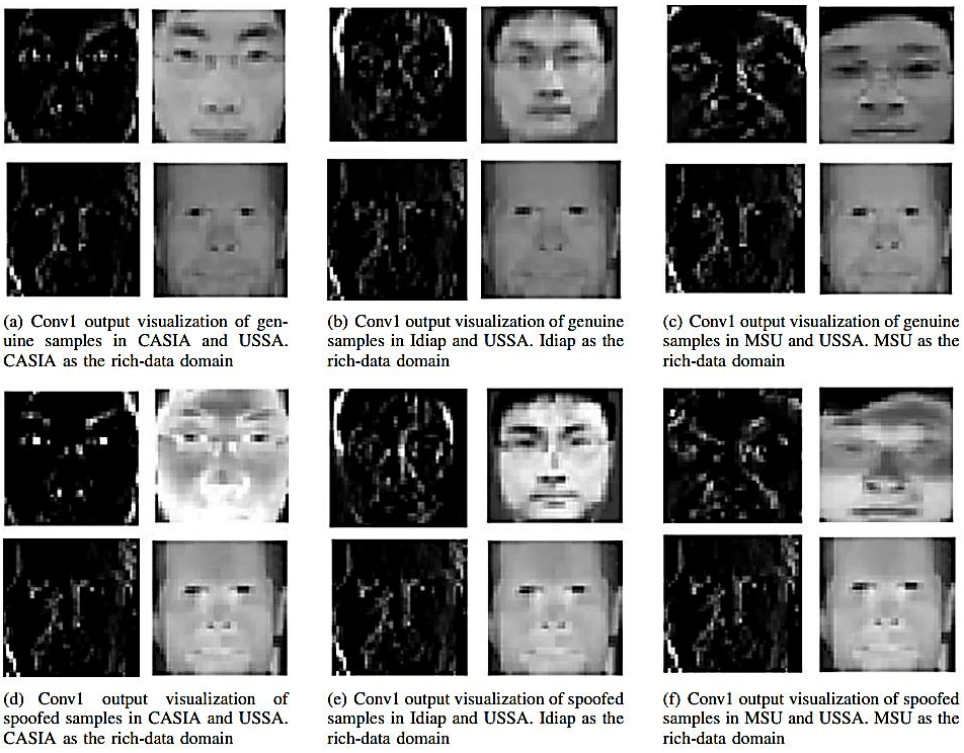



(c) MSU as the rich-data domain

**FIG 6:** Using Ussa As The Application-Specific Domain, We Compare The Same Error Rate Across Varying Topic Sizes.

❖ **Visualization**

We are interested in both making an objective comparison with the approach that was used as a benchmark and the information that we obtained while using the distillation framework that was supplied to us. The phoney face samples that are sent to the student network come from the CASIA, Idiap, and MSU databases, all of which were created from the USSA database. On the other hand, the instructor network accesses the USSA database in order to get both genuine and fabricated face samples. These databases may include both real and fabricated versions of the same face. Before carrying on with the work that was started in [62], we begin by generating a visual representation of the findings produced by the Conv1 layer. The component portions of the spatial feature map have been normalised to fall within the range. The method that we used to choose two of the sixty-four feature maps from CASIA, Idiap, and MSU, in addition to the USSA that accompanied them, is shown in Figure 6. In Figure 6, the Conv1 output of the genuine (or false) sample that was retrieved from the USSA database is shown in the first row of each subfigure. This information can be found in Figure 6. In the second row of each subfigure, the Conv1 result for a sample obtained from one of the rich-data domains (the CASIA, Idiap, or MSU databases) is shown. The sample may have been genuine or it could have been manufactured. The "Conv1 output" refers to the feature map that was selected as the output of the first convolutional layer [7-9].



(a) Conv1 output visualization of genuine samples in CASIA and USSA. CASIA as the rich-data domain

(b) Conv1 output visualization of genuine samples in Idiap and USSA. Idiap as the rich-data domain

(c) Conv1 output visualization of genuine samples in MSU and USSA. MSU as the rich-data domain

(d) Conv1 output visualization of spoofed samples in CASIA and USSA. CASIA as the rich-data domain

(e) Conv1 output visualization of spoofed samples in Idiap and USSA. Idiap as the rich-data domain

(f) Conv1 output visualization of spoofed samples in MSU and USSA. MSU as the rich-data domain

**FIG 6:** Visualisation Of The First Convolutional Layer's Selected Output Feature Maps (Conv1 Output) Using The Rich-Data Domains Casia, Idiap, And Msu And The Application-Specific Domain Ussa.

To begin, it should be mentioned that the findings from USSA's instructor networks, which were trained using a wide variety of datasets, are astonishingly similar. This indicates that the distillation paradigm is capable of explaining not just the possible similarities between teacher and student networks but also the basic aspects of the spoofing information used by USSA. This observation demonstrates that using several databases as the rich-data domains led to error rates that were similar to one another, which is consistent with the findings that are shown in Table I. Second, there is structural information present in both the teacher and the student networks. This information includes things like edges, textures, and human traits. It's possible that this type of information is held by both networks. It makes perfect sense that the teacher network would have a more in-depth knowledge of human appearance than the student network given that the teaching method used by the teacher network is distillation. Since only data that is relevant to spoofing is supplied to the student network, it is very improbable that information about a person's appearance would be valuable for cross-domain face anti-spoofing. It is common for the feature maps of the authentic samples to be darker than those of the false ones. Based on these findings, it seems that both teacher networks and student networks collaborate to maintain the lighting and colour information necessary for face ant spoofing. The findings

of the inquiry are consistent with the findings of this comment [10-12].

❖ **High Accuracy Achieved:**

The high degree of precision achieved in the detection of face spoofing attempts is one of the important outcomes that may be gained as a consequence of the technique. When integrated with concepts pertaining to the internet of things, the trained ANN model consistently demonstrated a remarkable ability to differentiate between actual and artificial faces. This high degree of accuracy is particularly crucial in applications such as access control systems and mobile device authentication, both of which place an emphasis on security and dependability.

## 7. Conclusion

Over the course of the last several years, it has grown more difficult to recognise a face parody. The vulnerability of face biometric systems to spoofing attacks has not been addressed despite the enormous research that has been conducted to find solutions that are dependable. Despite the recent spike in popularity that this work has seen, it is not yet completely under control at this point. In the context of solving challenges related to face recognition, we believe that residual training of deep neural networks has a much more encouraging future. On the other hand, due to the fact that it delivers such accurate outcomes, we have proposed that ANN be employed for picture recognition throughout this investigation. After using the ANN model, we have researched and evaluated a range of data augmentation methodologies. In addition, the system that is being investigated makes use of VGG-16 in order to learn how to classify features. Since ANN is dependent on layers that are implanted into neurons for communication, the layers need to be piled on top of one another so that the performance of this model can be accurately evaluated. Due to the fact that our solution is constructed on the knowledge distillation concept, it is in a position to successfully combine two distinct domains. It has been shown that using this approach is superior than either beginning from scratch to train a face anti-spoofing network or making use of transfer learning when attempting to solve the issue of a limited sample size that arises in an application-specific domain. Specifically, this issue pertains to the identification of fake faces. This is because of the way that the problem is presented. On the other hand, alternative PAD solutions require either the construction of a deep model from the ground up or the optimisation of an existing model.

## References


[1] J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proen¸ca, and J. Fierrez, "Ganprintr: Improved fakes and evaluation of the state of the art in face manipulation detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 1038–1048, 2020.

[2] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "*The journal of machine learning research*," vol. 15, pp. 1943–1955, 2014.

[3] X. Zhang, J. Zou, K. He, and J. Sun, "Accelerating very deep convolutional networks for classification and detection." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 1943–1955, 2016. [Online]. Available: http://dblp. uni-trier.de/db/journals/pami/pami38.html#ZhangZHS16 .

[4] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung, "Integration of image quality and motion cues for face antispoofing: *A neural network approach," Journal of Visual Communication and Image Representation*, vol. 38, pp. 451–460, 2016.

[5] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Scholkopf, and A. Smola, ¨ "A kernel two-sample test," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723–773, 2012.

[6] I. Chingovska, A. Anjos, S. Marcel, On the effectiveness of local binary patterns in face anti-spoofing, in: Proc. *International Conference of the Biometrics Special Interest Group*, Darmstadt, Germany, 2012, pp. 1–7.

[7] Shrivastava, A., Chakkaravarthy, M., Shah, M.A..A Novel Approach Using Learning Algorithm for Parkinson's Disease Detection with Handwritten Sketches. In Cybernetics and Systems, 2022

[8] Shrivastava, A., Chakkaravarthy, M., Shah, M.A., A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics. In *Healthcare Analytics*, 2023, 4, 100219

[9] Shrivastava, A., Chakkaravarthy, M., Shah, M.A.,Health Monitoring based Cognitive IoT using Fast Machine Learning Technique. In *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(6s), pp. 720–729

[10] *Shrivastava, A.*, *Rajput, N.*, *Rajesh, P.*, *Swarnalatha, S.R.*, IoT-Based Label Distribution Learning Mechanism for Autism Spectrum Disorder for Healthcare Application. In Practical Artificial Intelligence for Internet of Medical Things: Emerging Trends, Issues, and Challenges, 2023, pp. 305–321

[11] Boina, R., Ganage, D., Chincholkar, Y.D., .Chinthamu, N., Shrivastava, A., Enhancing Intelligence Diagnostic Accuracy Based on Machine Learning Disease Classification. In *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(6s), pp. 765–774

[12] Shrivastava, A., Pundir, S., Sharma, A., ...Kumar, R., Khan, A.K. Control of A Virtual System with Hand Gestures. In *Proceedings - 2023 3rd International Conference on Pervasive Computing and Social Networking, ICPCSN 2023,* 2023, pp. 1716–1721

[13] A. Benlamoudi, D. Samai, A. Ouafi, S. E. Bekhouche, A. Taleb-Ahmed, and A. Hadid, "Face spoofing detection using local binary patterns and fisher score," in 2015 3rd *International Conference on Control, Engineering Information Technology* (CEIT), 2015, pp. 1–5. doi: 10.1109/CEIT.2015.7233145.

[14] M. Asim, Z. Ming, and M. Y. Javed, "ANN based spatio-temporal feature extraction for face anti-spoofing," 2017 2nd *International Conference on Image, Vision and Computing* (ICIVC), pp. 234–238, 2017.

[15] M. Asim, Z. Ming, and M. Y. Javed, "ANN based spatio-temporal feature extraction for face anti-spoofing," in 2017 2nd *International Conference on Image, Vision and Computing (ICIVC)*, IEEE, 2017, pp. 234–238.

[16] A. Benlamoudi, D. Samai, A. Ouafi, S. E. Bekhouche, A. Taleb-Ahmed, and A. Hadid, "Face spoofing detection using local binary patterns and fisher score," in 2015 3rd *International Conference on Control, Engineering & Information Technology (CEIT)*, IEEE, 2015, pp. 1–5.