Scientific Research Publishing

Journal of Computer and Communications, 2024, 12, 86-101
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Explainable Deep Fake Framework for Images Creation and Classification

Majed M. Alwateer

Department of Computer Science, College of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia
Email: mwateer@taibahu.edu.sa

How to cite this paper: Alwateer, M.M. (2024) Explainable Deep Fake Framework for Images Creation and Classification. *Journal of Computer and Communications*, **12**, 86-101.
https://doi.org/10.4236/jcc.2024.125006

Received: April 17, 2024
Accepted: May 20, 2024
Published: May 23, 2024

Abstract

Deep learning is a practical and efficient technique that has been used extensively in many domains. Using deep learning technology, deepfakes create fake images of a person that people cannot distinguish from the real one. Recently, many researchers have focused on understanding how deepkakes work and detecting using deep learning approaches. This paper introduces an explainable deepfake framework for images creation and classification. The framework consists of three main parts: the first approach is called Instant ID which is used to create deepfacke images from the original one; the second approach called Xception classifies the real and deepfake images; the third approach called Local Interpretable Model (LIME) provides a method for interpreting the predictions of any machine learning model in a local and interpretable manner. Our study proposes deepfake approach that achieves 100% precision and 100% accuracy for deepfake creation and classification. Furthermore, the results highlight the superior performance of the proposed model in deep fake creation and classification.

Keywords

Deepfakes, Machine Learning, Deep Learning Fake Detection, Social Media LIME Technique

1. Introduction

Deep learning is a practical and efficient technique that has been widely applied in various domains, including computer vision and NLP. It has revolutionized these fields by enabling machines to learn and make predictions from large amounts of data [1]. Deepfake becomes harder to distinguish between actual and fake information. Despite the development of numerous deepfake detection and classification techniques, these systems frequently fail to identify deepfakes in practical scenarios. Especially when images are altered with new approaches not included in the training set, these systems frequently fail to differentiate images correctly. By using face database [2] face recognition technology is used to verify a person's identity. Biometric security includes face recognition among its categories. Deep learning-based networks are used by face recognition technology [3] [4] to recognize and learn certain face patterns. A mathematical representation is created using the face-related data.

In recent years, deep learning-based techniques have been established to identify and categorize deepfake photos, and numerous studies have been conducted to better understand how deepfakes operate [5] [6]. The study's findings suggest that the Xception model, a convolutional neural network (CNN) [7] architecture named "Xception: Deep Learning with Depthwise Separable Convolutions" by François Chollet in 2017, performs best when dealing with datasets that have fewer elements and manipulation techniques because it appears to be better at storing particular anomalies. On the other hand, when taught with a wider variety of datasets, the Vision Transformer performs better. Deepfake generation is carried out using generic adversarial networks based on artificial intelligence [8].

This paper introduces three models: the first approach is called Instant ID which uses ID embedding, a way of keeping the identity of the reference image while letting it change styles easily; the second approach called Deep Learning with Depthwise Separable Convolutions (Xception) classifies the real and deepfake images; the Third approach called Local Interpretable Model-Agnostic Explanations(LIME) provides a method for interpreting the predictions of any machine learning model in a local and interpretable manner. It approximates the model locally around the prediction of interest using a simpler, interpretable model, such as a linear model or decision tree.

The study will contribute to the understanding of deep learning techniques [9] [10] [11] and their application in addressing the challenges posed by deepfakes, ultimately aiming to improve the ability to distinguish between real and manipulated content in images.

The outlines of the research work include:

- A model called Xception is introduced to classify patterns and predict them from the real one.
- The Xception model is designed to achieve high accuracy on image classification tasks while being computationally efficient and requiring fewer parameters compared to other popular CNN models such as VGGNet and ResNe.
- The classification technique of the collected dataset is shown to be more accurate and predictive with 100% than other deep fake state-of-the-art studies.
- The LIME technique is used to visually interpret individual predictions generated by the model, emphasize important features, and provide explanations for the model's predictions.

2. Related Work

Convolutional Neural Networks (CNNs) [12] [13] appear to be better at storing certain abnormalities and perform well when dealing with datasets that include fewer elements and manipulation techniques, according to an investigation of many deep learning architectures. On the other hand, training the Vision Transformer with a wider variety of datasets increases its effectiveness.

Rafique *et al.* [14] proposed a framework that combines error-level analysis and deep learning for deep fake detection and classification. The framework involves performing error-level analysis on the image to determine if it has been modified, followed by deep feature extraction using Convolutional Neural Networks (CNNs). The extracted features are then classified using Support Vector Machines (SVMs) and K-Nearest Neighbors (KNN) algorithms.

Sugant *et al.* [15] focused on deep fake face recognition using deep learning techniques. It implements deep fake face image analysis using the Fisherface algorithm and Local Binary Pattern Histogram (FF-LBPH). The proposed model includes the use of CNNs for deep fake detection and classification.

Silva *et al.* [16] proposed a hierarchical interpreting forensics algorithm that incorporates humans in the detection loop. The work curates data through a deep learning detection algorithm and shares an explainable decision with humans alongside forensic analyses on the decision region.

In other articles, the variable analysis (VA) method is applied that identifies a small number of features for robust deep fake detection and classification [17]-[21]. Decision trees (DT) and logistic regression (LR) are used to illustrate the efficacy of the suggested model. Their study's dataset was obtained from UCI and Kaggle, and the findings showed that logistic regression performed better than other classifiers.

Today, deep learning (DL) [22] [23] [24] algorithms play a significant role in deepfake detection and classification. In several studies, it has been shown that DL is more effective in classifying ADS than ML [25] [26] [27]. Raj *et al.* [28] demonstrated that deep learning methods, specifically Convolutional Neural Networks (CNN), outperformed traditional ML methods for deepfake detection. The LSTM-RNN model is proposed for automated deepfake detection [29] [30].

The field of image forensics develops techniques to detect manipulated images. This comprehensive review covers state-of-the-art methods and datasets, benefiting researchers in this field. These insights provide a glimpse into the diverse and evolving landscape of deepfake detection and classification using deep learning and explainable techniques [31] [32] [33] [34].

The main drawbacks of the previous research works were less accurate classification in most cases. Our work creates a Deapfake image from the original one; and classifies the real and deepfake images. Additionally, provides a method for interpreting the predictions of any machine learning model in a local and interpretable manner.

3. Methodology

The purpose of this research is to introduce an explainable Deep Fake framework for Images creation and classification. The framework consists of three main parts: the first approach is called Instant ID which is used to create Deapfake image from the original one; the second approach called Deep Learning (Xception) classifies the real and deepfake images; the third approach called LIME provides a method for interpreting the predictions of any machine learning model in a local and interpretable manner.

Figure 1 explains the main three steps of the proposed model.

3.1. Data Preparation

These steps include two main phases: 1) Data Pre-processing and 2) Data Splitting. In the following paragraphs, each phase will be described in detail:

- **Data pre-processing:** This step involves identifying and handling any inconsistencies, errors, or missing values present in the collected data. Techniques such as removing duplicates, imputing missing values, or correcting inconsistencies will be employed to ensure data integrity. Moreover, this step involves transforming the data to ensure compatibility with the chosen AI model. In this step, category variables are encoded and numerical features are scaled. Before using a deep learning model, it is essential to resize data for fixing image size; and normalize the data because different attributes have different scales and values. All attribute data was normalized in the range [−1, 1] using the *Z* normalization approach, which removes the mean and scales the data to unit variance, as represented in Equation 1:

$$\text{Normalized Value} = \frac{X - \text{Mean}}{\text{Standard Deviation}} \tag{1}$$

- **Data splitting:** The pre-processed dataset is divided into subsets for training, testing, and validation. Common approaches include random splitting or stratified sampling to ensure representative subsets for each phase. The dataset was split into training and testing sets, with a ratio of approximately 70% training data and 30% testing data.

3.2. Data Creation

These steps include three main phases: 1) Data Augmentation, 2) Data Acquisition, and 3) Instant ID. In the following paragraphs, each phase will be described in detail:

- **Data augmentation** creating new data from old data, a technique known as "data augmentation" can be used to artificially expand the size of a training dataset. This lessens the chance of overfitting and enhances a model's capacity for generalization. The Keras Sequential API-based data augmentation pipeline. The pipeline consists of two data augmentation layers: 1) RandomFlip ("horizontal"): This layer randomly flips the input images horizontally. 2) Random Rotation (0.1): This layer randomly rotates the input images by up to 0.1 radians. By applying these transformations to the training data, the pipeline generates new images that are slightly different from the original images. Figure 2 shows a sample of data augmentation.
- **Data acquisition** is the process of sampling signals to measure actual physical

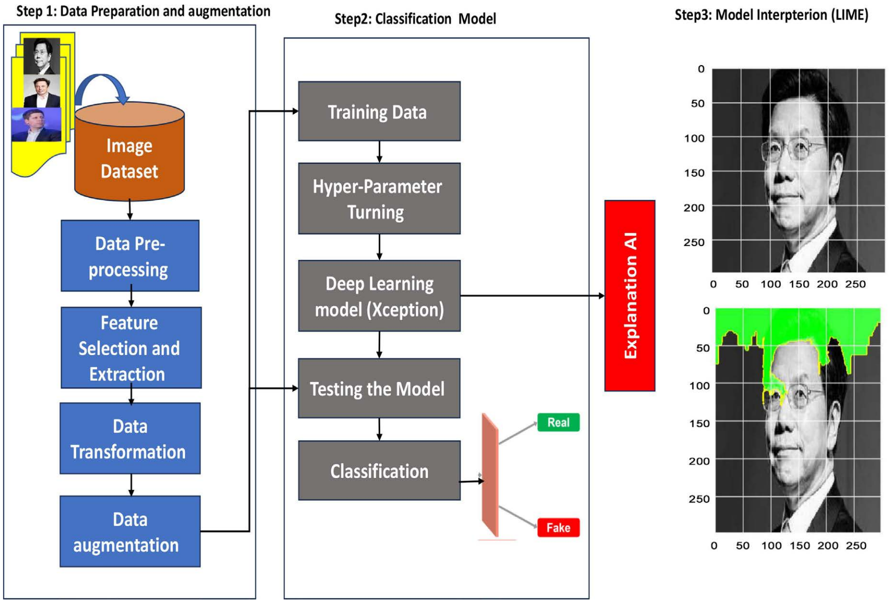


Figure 1. Main three steps for proposed model.

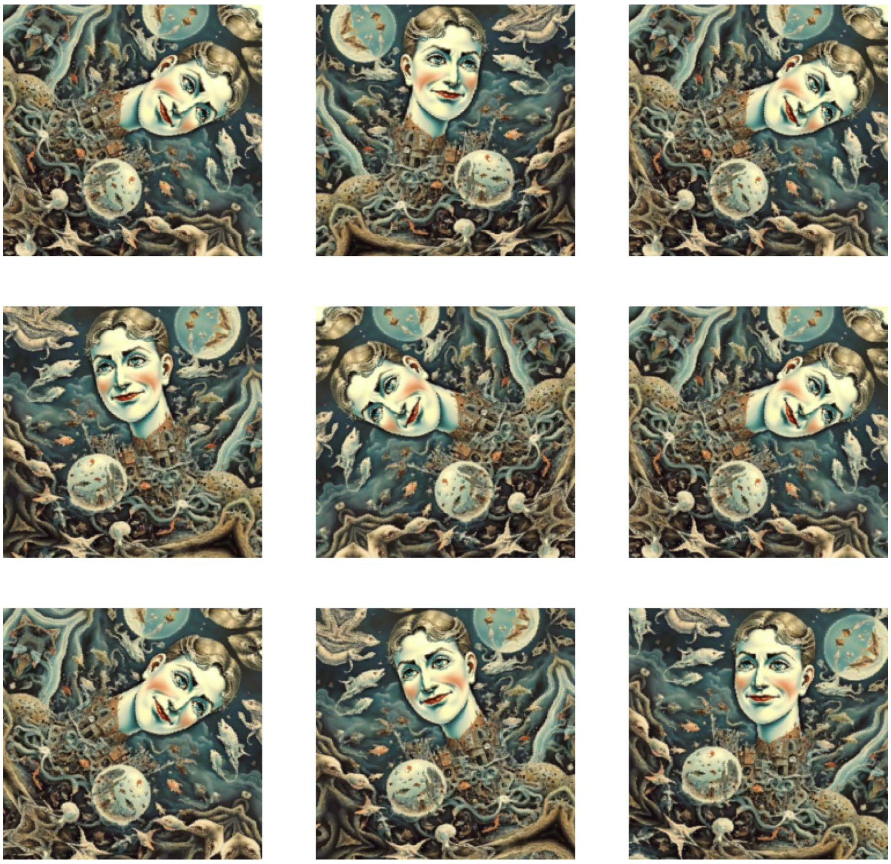


Figure 2. Data augmentation.

circumstances and translating the results into digital numerical values that a computer can control. It involves gathering data from various sensors or transducers and converting physical or electrical signals into digital data for further analysis and processing.

- **Instant ID** an innovative model in the field of identity generation, is revolutionizing the way we create and preserve identities. With its ability to produce high-fidelity images of individuals without any prior training data, Instant ID offers a zero-shot approach to identity generation. Moreover, Instant ID: Identity Preservation in Seconds with Zero Shots. With just one image, Instant ID is a brand-new, cutting-edge technique that achieves ID-Preserving generation without the need for tuning and supports a variety of downstream activities.

However, there are some potential factors that may obscure the replicability of the Instant ID model. These factors include:

Limited information on the model's implementation: While the Instant ID model is described in research papers and code repositories, there may be limited information available on the specific implementation details and training procedures. This lack of detailed documentation may make it challenging for researchers to replicate the model accurately.

Dependency on pre-trained text-to-image diffusion models: Instant ID is designed to seamlessly integrate with popular pre-trained text-to-image diffusion models like SD1.5 and SDXL. The replicability of Instant ID may depend on the availability and compatibility of these pre-trained models.

Lack of user responsibility: The developers of Instant ID emphasize that users are granted the freedom to create images using this tool but are obligated to comply with local laws and utilize it responsibly. The developers do not assume any responsibility for potential misuse by users. This lack of control over how the model is used may affect its replicability in certain contexts.

It is important to note that these potential factors may not necessarily hinder the replicability of the Instant ID model in all cases. Researchers and practitioners interested in replicating the model should refer to the available documentation, research papers, and code repositories for more information on the implementation and usage of Instant ID [35].

3.3. Xception Model

Convolutional neural network (CNN) architecture known as the Xception model was first presented by François Chollet in the 2017 publication "Xception: Deep Learning with Depthwise Separable Convolutions". Compared to other well-known CNN models like VGGNet and ResNet, it is intended to produce high accuracy on image classification tasks while being computationally efficient and using fewer parameters.

The Xception model consists of 36 convolutional layers, organized into 14 modules. Each module contains a depth wise separable convolution layer, Resize to the fix the length, a batch normalization layer, and an activation layer. The model also includes max pooling layers for down sampling and fully connected layers for classification.

The Xception model has been shown to achieve state-of-the-art results on various image classification datasets, including the ImageNet dataset. It is particularly well-suited for tasks where computational efficiency and low memory usage are important, such as mobile and embedded applications.

The model starts with a Rescaling layer that scales the pixel values of the input images to the range [0, 1]. The next layer is a Conv2D layer with 128 filters, a kernel size of 3, and a stride of 2. This layer applies convolution operations to the input images, extracting features from them. The output of this layer is then passed through a Batch Normalization layer and an Activation layer with the "relu" activation function.

The model then enters a series of residual blocks. Each residual block consists of the following sequence of layers: 1) Activation layer with the "relu" activation function. 2) SeparableConv2D layer with a specified number of filters (either 256, 512, or 728), a kernel size of 3, and a padding of "same". 3) Batch Normalization layer. 4) Activation layer with the "relu" activation function. 5) SeparableConv2D layer with the same number of filters as the previous layer, a kernel size of 3, and a padding of "same" Batch Normalization layer. 6) MaxPooling2D layer with a pool size of 3 and a stride of 2, used for down sampling.

After the residual blocks, the model includes a SeparableConv2D layer with 1024 filters, a kernel size of 3, and a padding of "same". The output of this layer is then passed through a Batch Normalization layer and an Activation layer with the "relu" activation function.

The model then uses a GlobalAveragePooling2D layer to reduce the spatial dimensions of the feature map to a single vector.

Finally, the activation function for the output layer depends on the number of classes: "sigmoid" for binary classification (2 classes) and "softmax" for multi-class classification.

3.4. Model Interpretation

The XAI framework is utilized to identify Deepfake and provide meaningful interpretations of the outcomes generated by the model. The framework leverages advanced AI techniques to analyze and interpret complex data patterns associated with Deepfake.

During the training phase, the model learns to recognize patterns and relationships within the data that are indicative of Deepfake. This process involves optimizing the model's parameters in order to reduce the difference between its predictions and the actual labels of Deepfake. Once the model is trained, the XAI framework focuses on providing interpretability of the model's outcomes. This is achieved through various techniques designed to shed light on the decision-making process of the model and the factors driving its predictions.

LIME (Local Interpretable Model-Agnostic Explanations) offers justifications for any classifier or regressor's predictions. LIME creates a fresh dataset of disturbances around the instance that needs to be explained. Then, each instance in the newly created dataset is classified using the machine learning classifier that has been trained. LIME has many advantages as: 1) Functions well with text, images, and tabular data Because of its versatility, 2) LIME can be used with a wide range of data formats, including text, pictures, and tabular data, 3) LIME is independent of the intricacies of the underlying model.

In this study, the Lime library provides a method for interpreting the predictions of any machine learning model in a local and interpretable manner. It approximates the model locally around the prediction of interest using a simpler, interpretable model, such as a linear model or decision tree.

4. Experimental Results

4.1. Dataset Specification and Collection

The dataset consists of a total of 589 files, which are divided into two distinct classes. Among these files, 472 are allocated for training purposes, while the remaining 117 are used for validation. The table, labeled as "Data set counting 2 classes," provides a comprehensive breakdown of the dataset. It includes information on the number of images in each category, namely Fake (0) and Real (1), for both the training and validation sets. The training set comprises 472 images, with 311 falling under the Fake category and 161 under the Real category. On the other hand, the validation set consists of 117 images, with 69 classified as Fake and 48 as Real. Overall, the dataset contains a total of 380 Fake images and 209 Real images as shown in **Table 1**.

Table 2 presents the performance evaluation metrics for the Xception model. The performance of the model is assessed on both the training and validation datasets. The table includes four metrics: accuracy, precision, recall, and F1-score. The model achieved perfect scores (1.00) for all metrics on both the training and validation datasets. This indicates that the model is highly effective in classifying the data and making accurate predictions.

Table 3 presents the classification report for the Xception model. The report includes four metrics: precision, recall, F1-score, and support. The model achieved perfect scores (1.00) for all metrics for both classes (0 and 1). This indicates that the model is highly effective in classifying the data and making accurate predictions. The accuracy of the model is also 1.00, which means that it correctly classified all 117 samples in the dataset (**Figure 3**).

4.2. Performance Evaluation

The proposed model's performance should be evaluated using classification metrics. Although classification accuracy is a commonly used metric, it may not be the most appropriate one for imbalanced datasets, where one class is much more represented than the others. Therefore, several other performance metrics have been developed, including precision, recall (also known as sensitivity) and F1-score. The confusion matrix in **Figure 4** presents the performance of Xception model. The matrix shows the number of false positives (FP), false negatives (FN), true negatives (TN) and true positives (TP), for each class. The model is classifying between the classes "Real" and "Fake". The TP value for the "Real"

Table 1. Data set counting 2 classes.

Data	#images	Fake (0)	Real (1)
Training	472	311	161
Validation	117	69	48
Total	589	380	209

Table 2. Performance of Xception model.

Metrics	Precision	Recall	Accuracy	F1-Score
Training	1.00	1.00	1.00	1.00
Validation	1.00	1.00	1.00	1.00

Table 3. Xception model classification report.

	Precision	Recall	F1-Score	Support
0	1.00	1.00	1.00	69
1	1.00	1.00	1.00	48
Accuracy			1.00	117
Macro avg	1.00	1.00	1.00	117
Weighted avg	1.00	1.00	1.00	117

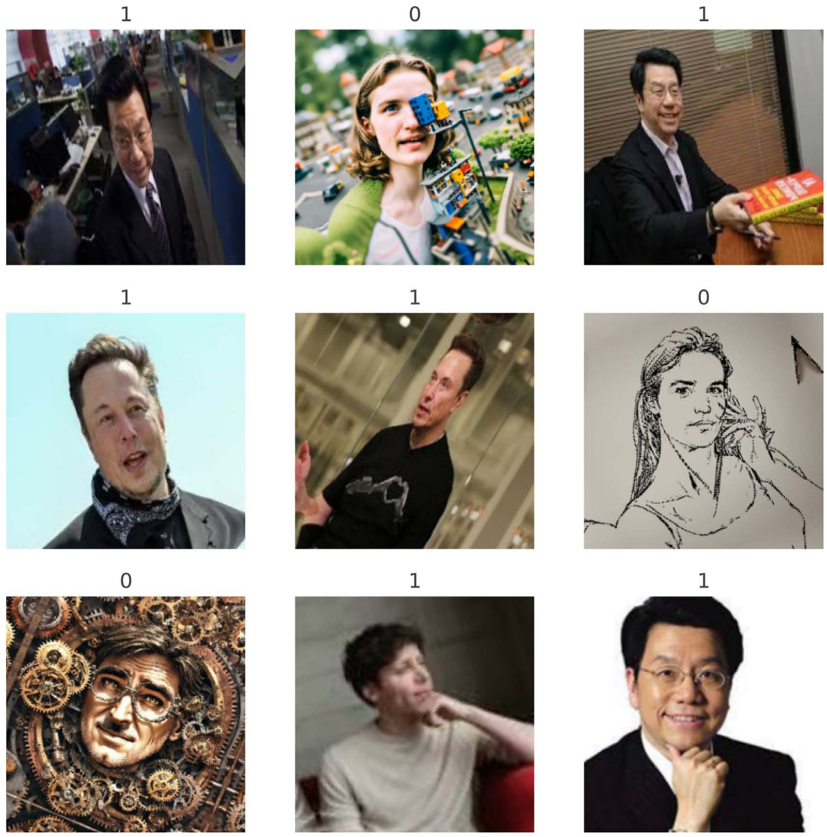


Figure 3. Data examples.

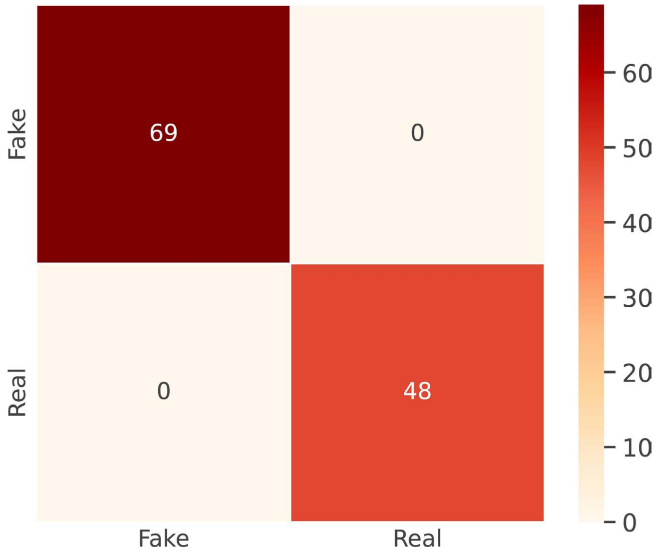


Figure 4. Confusion matrix of Xception model.

class is 69, which means that the model correctly classified 69 real samples as real. The FP value for the "Real" class is 0, which means that the model did not incorrectly classify any fake samples as real. The FN value for the "Real" class is 0, which means that the model did not incorrectly classify any real samples as fake. The TN value for the "Real" class is 48, which means that the model correctly classified 48 fake samples as fake [30] [36].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

Figure 5 shows the training and validation accuracy of a deep learning model. The model's accuracy gradually increases with the number of epochs on the training set, while the validation accuracy plateaus after around 50 epochs. To improve the model's performance, one could try reducing the number of epochs or using a regularization technique such as dropout.

Figure 6 shows the result of Xception model. The captions indicate the predicted probability of the image being fake, the true probability of the image being fake, and whether the image is actually fake or real. The model is able to correctly classify all of the real samples and all of the fake samples. This indicates that the model is highly effective in making accurate predictions.

To explain a prediction for a particular instance, LIME generates a set of synthetic data points by perturbing the original instance and observing how the model's prediction changes. The weights of the synthetic data points are then used to determine the importance of each feature in the original instance for the model's prediction.

LIME is particularly useful for explaining complex machine learning models, such as deep neural networks, which can be difficult to interpret directly. By

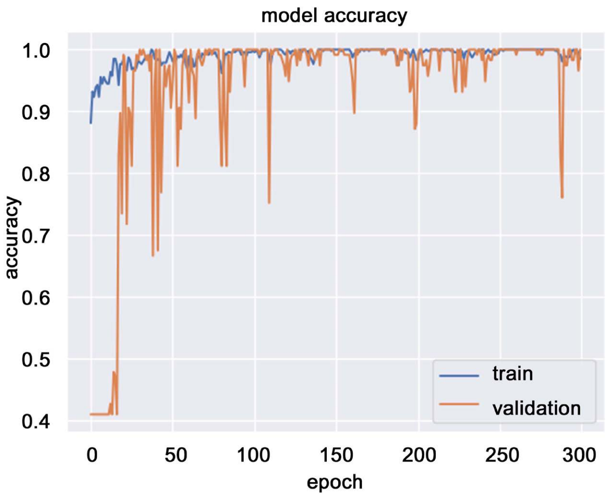


Figure 5. Accuracy curve of Xception model.

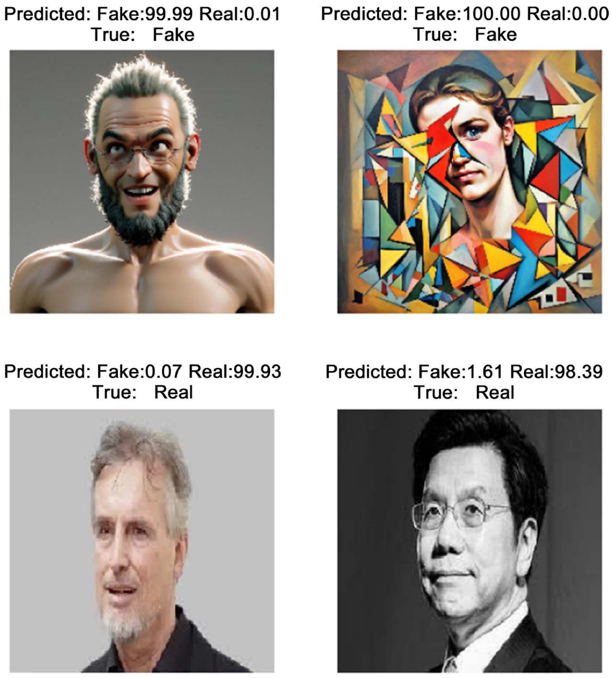

Predicted: Fake:99.99 Real:0.01
True: Fake
Predicted: Fake:100.00 Real:0.00
True: Fake
Predicted: Fake:0.07 Real:99.93
True: Real
Predicted: Fake:1.61 Real:98.39
True: Real

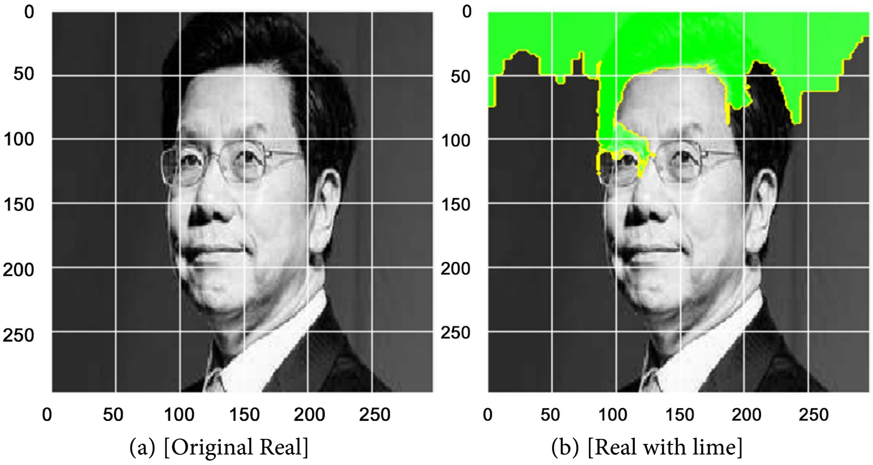
Figure 6. The result of Xception model.

approximating the model locally, LIME provides a simplified explanation that is easier to understand and interpret.

Figure 7 shows the explanation for the two classes using the LIME library. From Figure 6, the explanation displays the top 10 superpixels that contribute positively to the real class, while concealing the rest of the image. This insightful visualization helps identify the specific visual features the model relies on for its prediction, shedding light on the decision-making process.

4.3. Comparison with Other Models

Table 4 presents a comparative analysis of deep fake detection approaches. The



(a) [Original Real] (b) [Real with lime]

Figure 7. The explanation for the two classes.

Table 4. The comparative analysis of deep fake detection approaches.

Approach	Accuracy	F1
The proposed model	100	100
VGG16 and CNN models [25]	94	94
Multimodal network [15]	61	61

evaluated approaches include the proposed model and the VGG16 and CNN architecture [37] [38]. The dataset used for evaluation comprises real and fake faces, as well as photoshopped real and fake faces. The accuracy and F1 scores are reported as performance metrics. The proposed model achieved a perfect accuracy and F1 score of 100, indicating its effectiveness in detecting deep fake images. In comparison, the VGG16 and CNN architecture achieved an accuracy and F1 score of 94, demonstrating relatively lower performance in identifying manipulated images. The results highlight the superior performance of the proposed model in deep fake creation and classification than the state-of-arts.

5. Conclusion

The proliferation of deep fake content in images, poses a significant challenge in the digital landscape. The ease of access to advanced tools and computing infrastructure has facilitated the creation and dissemination of deepfakes, leading to the potential spread of disinformation, hoaxes, and panic. As a result, the need for robust deepfake classification using deep learning techniques has become increasingly critical. This comprehensive study has delved into the various aspects of deep fake detection and classification, leveraging insights from state-of-the-art methods and datasets. The review has highlighted the evolving landscape of deepfake detection and classification, encompassing the utilization of Convolutional Neural Networks (CNNs) called Xception and other deep learning architectures for image analysis. Additionally, LIME provides a method for interpreting ML prediction for the classification image as real or fake. The research community's efforts in this domain, as evidenced by numerous papers and code repositories dedicated to deepfake detection, reflect the urgency and importance of addressing the challenges posed by deep fakes. As the use of deep learning techniques to manipulate images continues to grow, the need for effective deepfake detection systems becomes even more pronounced. In conclusion, this comprehensive study serves as a valuable resource for researchers and practitioners in the field of deepfake detection and classification. It seeks to support ongoing attempts to counter the spread of deepfake content and its possible negative effects on society by illuminating the most recent methods and difficulties. The shortcomings of the existing approaches demonstrate the continued need for the development of a reliable and effective deepfake detection and classification solution based on ML and DL techniques. Moreover, it is important to consider the diversity and representativeness of the datasets to ensure that the framework's performance is not limited to specific scenarios or domains. Future work could focus on using a big Image data to test the model well.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Zobaed, S., Rabby, F., Hossain, I., Hossain, E., Hasan, S., *et al.* (2021) DeepFakes: Detecting Forged and Synthetic Media Content Using Machine Learning. In: Montasari, R. and Jahankhani, H., Eds., *Artificial Intelligence in Cyber Security*: *Impact and Implications*, Springer, Berlin, 177-201. https://doi.org/10.1007/978-3-030-88040-8_7

[2] Lee, H., Park, S., Yoo, J., Jung, S. and Huh, J. (2020) Face Recognition at a Distance for a Stand-Alone Access Control System. *Sensors*, **20**, 785-796. https://doi.org/10.3390/s20030785

[3] Al Bdairi, A., Xiao, Z., Alkhayyat, A., Humaidi, A., Fadhel, M., Taher, B., Alzubaidi, L., Santamaria, J. and Al-Shamma, O. (2022) Face Recognition Based on Deep Learning and FPGA for Ethnicity Identification. *Applied Sciences*, **10**, 2605-2520. https://doi.org/10.3390/app12052605

[4] Sarangi, P., Nayak, D., Panda, M. and Majhi, B. (2022) A Feature-Level Fusion Based Improved Multimodal Biometric Recognition System Using Ear and Profile Face. *Journal of Ambient Intelligence and Humanized Computing*, **13**, 1867-1898. https://doi.org/10.1007/s12652-021-02952-0

[5] Mahmud, M., Kaiser, M.S., Rahman, M.A., Wadhera, T., Brown, D.J., Shopland, N., Burton, A., Hughes-Roberts, T., Mamun, S.A., Ieracitano, C. and Tania, M.H. (2022) Towards Explainable and Privacy-Preserving Artificial Intelligence for Personalisation in Autism Spectrum Disorder. *International Conference on Human-Computer Interaction*, Gothenburg, Sweden, 26 June-1 July 2022, 356-370. https://doi.org/10.1007/978-3-031-05039-8_26

[6] Ali, S., Akhlaq, F., Imran, A.S., Kastrati, Z., Daudpota, S.M. and Moosa, M. (2023) The Enlightening Role of Explainable Artificial Intelligence in Medical & Healthcare Domains: A Systematic Literature Review. *Computers in Biology and Medicine*, **166**, Article 107555. https://doi.org/10.1016/j.compbiomed.2023.107555

[7] Yousfi, Y., Butora, J., Khvedchenya, E. and Fridrich, J. (2020) Imagenet Pre-Trained CNNs for JPEG Steganalysis. *Proceedings of the* 2020 *IEEE International Workshop on Information Forensics and Security* (*WIFS*), New York, 6-11 December 2020, 1-6. https://doi.org/10.1109/WIFS49906.2020.9360897

[8] Thambawita, V., Isaksen, J., Hicks, S., Ghouse, J., Ahlberg, G., Linneberg, A., Grarup, N., Ellervik, C. and Olesen, M. (2021) Deepfake Electrocardiograms Using Generative Adversarial Networks Are the Beginning of the End for Privacy Medicine. *Scientific Reports*, **11**, 21869-21889. https://doi.org/10.1038/s41598-021-01248-9

[9] El-Sayed Atlam, K.M., Fuketa, M. and Ich Aoe, J. (2003) Document Similarity Measurement Using Field Association Term. *Information Processing and Management Journal*, **39**, 809-824. https://doi.org/10.1016/S0306-4573(03)00019-0

[10] Farsi, M., Hosahalli, D., Manjunatha, B., Gad, I., Atlam, E.-S., Ahmed, A., *et al.* (2021) Parallel Genetic Algorithms for Optimizing the SARIMA Model for Better Forecasting of the NCDC Weather Data. *Alexandria Engineering Journal*, **60**, 1299-1316. https://doi.org/10.1016/j.aej.2020.10.052

[11] Noor, T., Almars, A., Alwateer, M., Almaliki, M., *et al.* (2022) Sarima: A Seasonal Autoregressive Integrated Moving Average Model for Crime Analysis in Saudi Arabia. *Electronics*, **11**, 3986-3998. https://doi.org/10.3390/electronics11233986

[12] Jung, T., Kim, S. and Kim, K. (2020) Deepvision: Deepfakes Detection Using Human Eye Blinking Pattern. *IEEE Access*, **8**, 83144-83154. https://doi.org/10.1109/ACCESS.2020.2988660

[13] Hsu, C., Zhuang, Y. and Lee, C. (2020) Deep Fake Image Detection Based on Pairwise Learning. *Applied Sciences*, **10**, 370-386. https://doi.org/10.3390/app10010370

[14] Rafique, R., Gantassi, R., Amin, R., Frnda, J., Mustapha, A. and Alshehri, A.H. (2023) Deep Fake Detection and Classification Using Error-Level Analysis and Deep Learning. *Scientific Reports*, **13**, Article No. 7422. https://doi.org/10.1038/s41598-023-34629-3

[15] Suganthi, S., Ayoobkhan, M.U.A., Bacanin, N., Venkatachalam, K., *et al.* (2022) Deep Learning Model for Deep Fake Face Recognition and Detection. *PeerJ Computer Science*, **8**, e881. https://doi.org/10.7717/peerj-cs.881

[16] Silva, S.H., Bethany, M., Votto, A.M., Scarff, I.H., Beebe, N. and Najafirad, P. (2022) Deepfake Forensics Analysis: An Explainable Hierarchical Ensemble of Weakly Supervised Models. *Forensic Science International*: *Synergy*, **4**, Article ID: 100217. https://doi.org/10.1016/j.fsisyn.2022.100217

[17] Thabtah, F. (2019) Machine Learning in Autistic Spectrum Disorder Behavioral Research: A Review and Ways Forward. *Informatics for Health and Social Care*, **44**, 278-297. https://doi.org/10.1080/17538157.2017.1399132

[18] Thabtah, F., Kamalov, F. and Rajab, K. (2018) A New Computational Intelligence Approach to Detect Autistic Features for Autism Screening. *International Journal of Medical Informatics*, **117**, 112-124.
https://www.sciencedirect.com/science/article/pii/s1386505618300546
https://doi.org/10.1016/j.ijmedinf.2018.06.009

[19] Hossain, S., Islam, M.A., Quinn, F., Huq, J.M. and Moni, M. (2019) Machine Learning and Bioinformatics Models to Identify Gene Expression Patterns of Ovarian Cancer Associated with Disease Progression and Mortality. *Journal of Biomedical Informatics*, **100**, 310-313. https://doi.org/10.1016/j.jbi.2019.103313

[20] Howlader, K., Satu, M., Barua, A. and Moni, M. (2018) Mining Significant Features of Diabetes Mellitus Applying Decision Trees: A Case Study in Bangladesh. https://doi.org/10.1101/481994

[21] Thabtah, F. (2017) Autism Spectrum Disorder Screening: Machine Learning Adaptation and Dsm-5 Fulfillment. *Proceedings of the* 1*st International Conference on Medical and Health Informatics*, Taichung City, 20-22 May 2017, 1-6. https://doi.org/10.1145/3107514.3107515

[22] Malki, Z., Atlam, E.-S., Hassanien, A.E., Dagnew, G., Elhosseini, M.A. and Gad, I. (2020) Association between Weather Data and COVID-19 Pandemic Predicting Mortality Rate: Machine Learning Approaches. *Chaos*, *Solitons and Fractals*, **138**, Article ID: 110137. https://doi.org/10.1016/j.chaos.2020.110137

[23] Malki, Z., Atlam, E.-S., Ewis, A., Dagnew, G., Reda, A., Elmarhomy, G., *et al.* (2020) ARIMA Models for Predicting the End of COVID-19 Pandemic and the Risk of a Second Rebound. *Journal of Neural Computing and Applications*, **33**, 2929-2948. https://doi.org/10.1007/s00521-020-05434-0

[24] Malki, Z., Atlam, E.-S., Ewis, A., Dagnew, G., Ghoneim, O.A., Mohamed, A.A., Abdel-Daim, M.M. and Gad, I. (2021) The Covid-19 Pandemic: Prediction Study Based on Machine Learning Model. *Journal of Environmental Science and Pollution Research*, **28**, 40496-40506. https://doi.org/10.1007/s11356-021-13824-7

[25] Almars, A.M., Alwateer, M., Qaraad, M., Amjad, S., Fathi, H., Kelany, A.K., Hussein, N.K. and Elhosseini, M. (2021) Brain Cancer Prediction Based on Novel Interpretable Ensemble Gene Selection Algorithm and Classifier. *Diagnostics*, **11**, Article No. 1936. https://doi.org/10.3390/diagnostics11101936

[26] Badawy, M., Almars, A.M., Balaha, H.M., Shehata, M., Qaraad, M. and Elhosseini, M. (2023) A Two-Stage Renal Disease Classification Based on Transfer Learning with Hyperparameters Optimization. *Frontiers in Medicine*, **10**, Article ID: 1106717. https://doi.org/10.3389/fmed.2023.1106717

[27] Alwateer, M., Almars, A.M., Areed, K.N., Elhosseini, M.A., Haikal, A.Y. and Badawy, M. (2021) Ambient Healthcare Approach with Hybrid Whale Optimization Algorithm and Naive Bayes Classifier. *Sensors*, **21**, Article No. 4579. https://doi.org/10.3390/s21134579

[28] Raj, S. and Masood, S. (2020) Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. *Procedia Computer Science*, **167**, 994-1004. https://doi.org/10.1016/j.procs.2020.03.399

[29] Kollias, K.F., Syriopoulou-Delli, C.K., Sarigiannidis, P. and Fragulis, G.F. (2021) The Contribution of Machine Learning and Eye-Tracking Technology in Autism Spectrum Disorder Research: A Review Study. 2021 10*th International Conference on Modern Circuits and Systems Technologies* (*MOCAST*). Thessaloniki, 05-07 July 2021, 1-4. https://doi.org/10.1109/MOCAST52088.2021.9493357

[30] Atlam, E., Ewis, A., Abd El-Raouf, M., Ghoneim, O. and Gad, I. (2022) A New Approach in Identifying the Psychological Impact of Covid-19 on University Student's Academic Performance. *Alexandria Engineering Journal*, **61**, 5223-5233. https://doi.org/10.1016/j.aej.2021.10.046

[31] Hooshmand, M.K., Huchaiah, M.D., Alzighaibi, A.R., Hashim, H., Atlam, E.-S. and Gad, I. (2024) Robust Network Anomaly Detection Using Ensemble Learning Approach and Explainable Artificial Intelligence (Xai). *Alexandria Engineering Journal*, **94**, 120-130. https://doi.org/10.1016/j.aej.2024.03.041

[32] Masud, M., Almars, A.M., Rokaya, M.B., Meshref, H., Gad, I. and Atlam, E.-S. (2024) A Novel Light-Weight Convolutional Neural Network Model to Predict Alzheimer's Disease Applying Weighted Loss Function. *Journal of Disability Research*, **3**, Article ID: 20240042. https://doi.org/10.57197/JDR-2024-0042

[33] Atlam, E.-S., Masud, M., Rokaya, M., Meshref, H., Gad, I. and Almars, A.M. (2024)

Easdm: Explainable Autism Spectrum Disorder Model Based on Deep Learning. *Journal of Disability Research*, **3**, Article ID: 20240003. https://doi.org/10.57197/JDR-2024-0003

[34] Noor, T.H., Almars, A.M., El-Sayed, A. and Noor, A. (2022) Deep Learning Model for Predicting Consumers' Interests of IoT Recommendation System. *International Journal of Advanced Computer Science and Applications*, **13**, 161-170. https://doi.org/10.14569/IJACSA.2022.0131022

[35] Wang, Q., Bai, X., Wang, H., Qin, Z. and Chen, A. (2024) Instantid: Zero-Shot Identity Preserving Generation in Seconds.

[36] Gad, I. and Hosahalli, D. (2020) A Comparative Study of Prediction and Classification Models on NCDC Weather Data. *International Journal of Computers and Applications*, **44**, 414-425. https://doi.org/10.1080/1206212X.2020.1766769

[37] Raza, A., Munir, K. and Almutairi, M. (2022) A Novel Deep Learning Approach for Deepfake Image Detection. *Applied Sciences*, **12**, Article No. 9820. https://doi.org/10.3390/app12199820

[38] Lewis, J.K., Toubal, I.E., Chen, H., *et al.* (2020) Deepfake Video Detection Based on Spatial, Spectral, and Temporal Inconsistencies Using Multimodal Deep Learning. 2020 *IEEE Applied Imagery Pattern Recognition Workshop* (*AIPR*), Washington DC, 13-15 October 2020, 1-9.