IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SECUREFACES: FACIAL AUTHENTICATION WITH DEEPFAKE DEFENSE

[1]Jayanthi M G, [2]Jai Surya R, [3]Bhoomika M P, [4]Danisha BN

[1]Professor, [2]Student, [3]Student, [4]Student

Department of Computer Science,

Cambridge Institute of Technology, Bengaluru, India

Abstract: SecureFaces stands out as a resolute guardian of online security in a time of digital interactions by fusing cutting-edge face authentication with a powerful Deepfake detecting tool. In this research, two powerful models—MesoNet for quick authentication and ResNetLSTM for increased accuracy—are shown. MesoNet and ResNetLSTM each address different user priorities. To ensure efficiency and data integrity, the system starts with an easy-to-use registration module that collects and safely stores facial data in MangoDB. Users can select between MesoNet and ResNetLSTM during the authentication process, depending on their unique needs, creating a customized identity verification strategy. SecureFaces is essentially an innovative cybersecurity solution that embodies trust, transparency, and adaptability in the digital sphere, going beyond simple facial authentication. This initiative acts as a beacon, guiding users through a trustworthy and safe authentication process as internet threats change.

Index Terms - Cybersecurity, Facial Authentication, Deepfake, Machine Learning, Deep Learning, Artificial Intelligence.

I. INTRODUCTION

The arrival of the digital age has fundamentally changed how we produce, distribute, and engage with multimedia material. While this digital transition has created a wealth of opportunities, it has also created brand-new, previously unknown obstacles. The spread of deepfake technology is one of these problems that is most noticeable. Deepfakes are artificial intelligence (AI)-generated or edited multimedia, such as videos, audio recordings, and photographs, that effectively mimic the appearance and voice of real people, frequently with the aim of deceiving or manipulating the user. Generative adversarial networks (GANs) are used by this technology to create content that frequently cannot be distinguished from real media. Deepfake technology has wide-ranging effects that touch on many parts of contemporary culture. While deepfakes have useful and inventive uses, such as in the entertainment sector, the possibility of harmful exploitation is an important concern.

Deepfakes can be used for a variety of malicious reasons, such as:

1. Identity Theft Prevention: Deepfakes can be used to impersonate people, which may result in fraud or identity theft. A system of authentication can confirm the authenticity of identification documents, images, or videos, preventing harmful individuals from using another person's identity to commit crimes.
2. Secure Online Transactions: Deepfakes represent a danger to financial transactions with the growth of online banking and e-commerce. By confirming that the identities of users are real, an authentication system can improve security and lower the risk of financial fraud.
3. Enhanced Privacy Protection: Deepfakes pose a danger to privacy since they can be used to modify sensitive or obscene content. Such media's integrity can be confirmed by an authentication system to stop unauthorised dissemination and modification.

Maintaining the trust that supports digital interactions is another reason for deepfake authentication. The capacity to confirm the validity of multimedia content becomes crucial in a society where information and media content play a crucial role in influencing public perception and decision-making. We hope to uphold the authenticity of media, news, and online transactions by using deepfake authentication. In doing so, we hope to create an environment where people can communicate with confidence, have faith in the accuracy of what they read, and feel secure in their digital interactions. Deepfake authentication is a compelling and essential technological achievement because it supports the greater goals of cybersecurity, privacy preservation, and the ongoing defense of the digital world against emerging threats.

The main goal is to create and execute a facial authentication system that is more advanced than current approaches. Use MesoNet as a fundamental element to provide instantaneous user authentication. As the first screening mechanism, this model's reputation for computing efficiency allows for quick and fast authentication procedures. To carry out in-depth analysis and detection of deepfake changes, incorporate ResNetLSTM into the system. This model contributes to increased security by improving the system's capacity to detect even complex deepfake efforts through its emphasis on temporal correlations. As the Presentation Layer, create a user-friendly website that acts as the login and registration interface for users. The aim is to provide a user-friendly and easily navigable platform that facilitates the registration process and offers clear communication of authentication results.

II. Related Works

Literature review was performed to understand the existing learning algorithms and to choose the suitable supervised and unsupervised method for image classification. As the study was made to compare supervised and unsupervised algorithms, the literature review was performed to identify the most effective algorithm of each kind. The algorithms identified were further used in experimentation.

The authors of [1] have put forth a GAN-based technique. Deepfakes have become more commonplace due to the ease of access to audio-visual content on social media, the availability of tools such as TensorFlow and Keras, open-source trained models, and reasonably priced computing infrastructure. These misrepresented media outlets are used for fraud, hoaxes, revenge porn, misinformation, and interference with government operations. Prior studies primarily focused on identifying deepfake photos and videos.

The authors of [2] have put forth a MesoNet Model for detection. To combat the rise of hyper-realistic manipulated videos, like Deepfake and Face2Face, this paper presents an automatic and effective method for detecting face tampering in videos. This method makes use of deep learning, in contrast to traditional image forensics, which has problems with video compression. The paper highlights mesoscopic image properties by presenting two networks, each with a restricted number of layers.

The FaceSwap method has been proposed by the authors in [3]. Recently, the public has had easy access to several face-manipulation techniques for videos, including FaceSwap and deepfake. With little effort, these techniques make it possible to easily edit faces in video sequences to look incredibly realistic. Although these tools have valid uses, when they are abused, they can result in major social problems like the propagation of false information and the manipulation of content for the purpose of cyberbullying.

The authors of [4] proposed a novel framework based on ResNet/LSTM combined model for short-term load forecasting. This research suggests a short-term load forecasting strategy based on the combined ResNet/LSTM model. Two steps make up the suggested model. First, latent features from weekly and daily load data are extracted by ResNet. The encoded feature vector is then trained with dynamics using LSTM, which makes the prediction appropriate for unstable load data. The suggested model benefits from the ability to forecast load data with both regularity and irregularity by utilising ResNet and LSTM.

III. Research Methodology

The primary objective of this project is to improve digital security by creating and deploying a facial authentication system with built-in Deepfake detection capabilities. Innovative solutions are required to protect identity verification procedures, online transactions, and sensitive data from the growing threat posed by deepfake technologies. The project intends to strengthen cybersecurity safeguards by leveraging cutting-edge models like ResNetLSTM for strong Deepfake detection and MesoNet for quick and real-time user verification.

The development, deployment, and maintenance of a thorough face authentication system with a particular emphasis on mitigating the risks associated with deepfake technology are all included in the project's scope. The project will integrate the MesoNet and ResNetLSTM models, taking advantage of their individual advantages to guarantee accuracy and efficiency in the authentication procedure. Real-time operation is also included in the project scope, which enables the system to respond to changing cybersecurity conditions and adjust to changing scenarios.

3.1 MesoNet Model

A deep learning architecture represented in Fig. 1, is created especially for deepfake detection is the MesoNet model. It has been demonstrated that this efficient and small model can produce state-of the-art outcomes for this task. There are three primary parts to the MesoNet architecture:

1. Convolutional Neural Network (CNN): This part of the system oversees taking the input data and extracting its spatial properties. It is made up of several convolutional layers, with a pooling layer sitting in between each one. The model is more effectively trained and deployed when the dimensionality of the data is reduced by the pooling layers. It has four convolutional layers, with a pooling layer in between each. While the pooling layers lower the dimensionality of the data, the convolutional layers learn to extract features from the input data. The kernel sizes of the final two convolutional layers are 5x5, while the kernel sizes of the first two convolutional layers are 3x3. As a result, the network may learn features at various scales.
2. Recurrent Neural Network (RNN): The temporal properties of the input data are extracted by the RNN component. There is just one RNN layer in it. To detect deepfakes, the RNN layer must learn to represent the temporal correlations between a video's frames. Because of this, it is a good fit for tasks like deepfake detection, in which it is crucial to understand the temporal relationships between video frames. The MesoNet model's RNN layer is set up with 16 units. This indicates that the layer's 16 hidden states correspond to its input data memory.
3. Fully connected layer: The CNN and RNN components extracted spatial and temporal information are combined by the fully connected layer. Whether the input data is a real or false video is shown by the likelihood score that is produced. It has been demonstrated that the MesoNet model, a potent deep learning architecture, is useful for deepfake detection. It is a small, effective model that is simple to use and train.

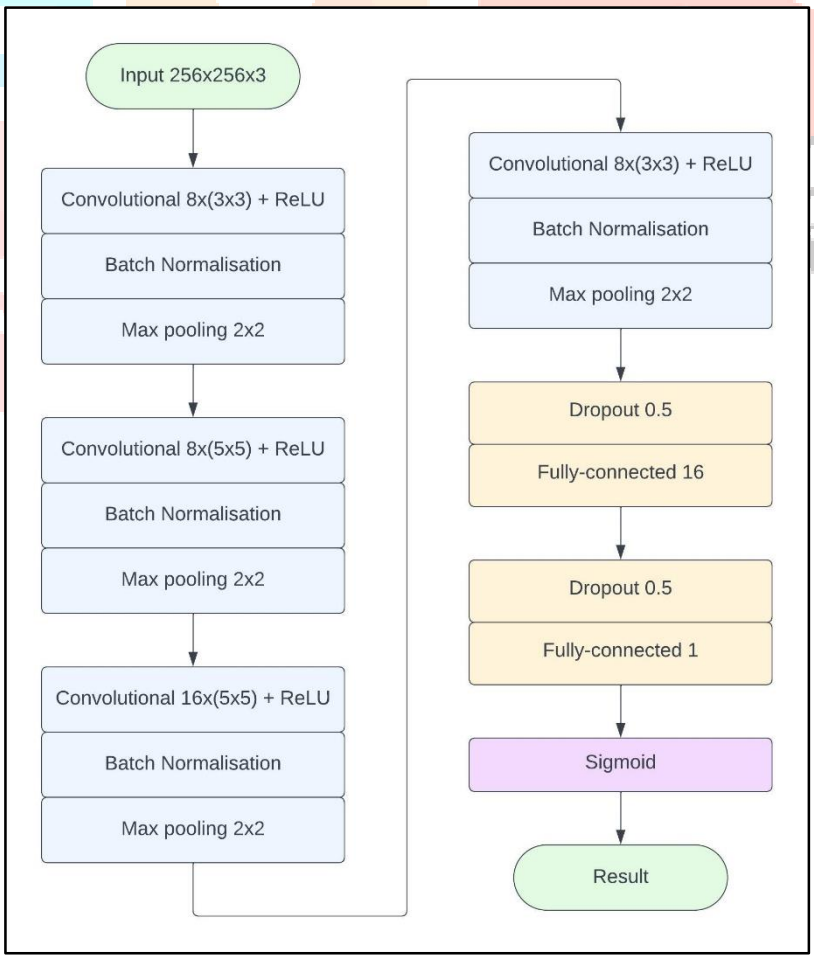

Fig.1: MesoNet Model Architecture

3.2 ResNetLSTM Model

Combining the advantages of long short-term memory (LSTM) networks and residual neural networks (ResNets), the ResNetLSTM model is a deep learning architecture represented in Fig. 2. Although it was first suggested for deepfake detection, it has subsequently been used for many other purposes, such as video analysis and picture categorization. There are three primary parts to the ResNetLSTM architecture:

1. Residual network (ResNet): This part of the system is in charge of taking the input data and extracting its spatial properties. It is made up of several residual blocks, with two convolutional layers and a shortcut connection in between each one. For tasks like deepfake detection, the network may learn long-range dependencies in the data thanks to the shortcut connection. A sequence of residual blocks, each consisting of two convolutional layers and a shortcut connection, make up the ResNet architecture. For tasks like deepfake detection, the network may learn long-term dependencies in the input information thanks to the shortcut connection. Typically, the ResNet part of the ResNetLSTM model is set up with eighteen residual blocks. The number of channels in the latter residual blocks is greater than that of the initial few residual blocks. As a result, the network may learn features at various scales.
2. Bidirectional LSTM network: This network oversees taking the input data and extracting temporal properties. It is composed of two Long Short-Term Memory (LSTM) layers, one for forward processing and one for backward processing. For applications like video analysis, this enables the network to understand temporal correlations in the data. A sequence of LSTM cells, each with an input gate, an output gate, and a forget gate, make up the LSTM architecture. The network is able to identify long-range dependencies in the data because the gates regulate the information flow within the cell. As a result, the network can discover temporal correlations in the information.

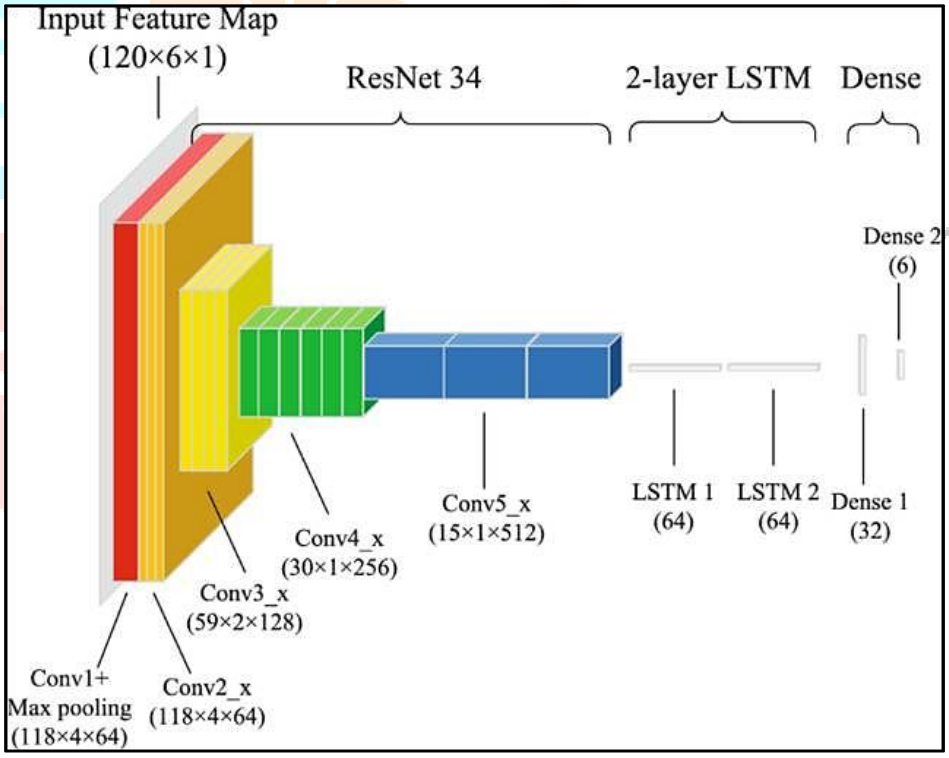

Fig. 2: ResNetLSTM Model Architecture

3. Fully connected layer: The ResNet and bidirectional LSTM network's extracted spatial and temporal information are combined by the fully connected layer. Whether the input data is a real image or video or a false one is shown by the likelihood score that is produced. Whether the input data is a real image or video or a false one is shown by the likelihood score that is produced. It has been demonstrated that the ResNetLSTM model is a potent deep learning architecture that performs well on a range of tasks. This paradigm is adaptable and can be used in many different contexts.

3.3 System Architecture

A system that detects deepfakes through deep learning is known as a deepfake authentication system. Deepfakes are artificial media that are produced by editing photos or videos to give the impression that someone is saying or acting in a way that they never would have. The architecture is shown in Fig. 3.

The following elements usually make up a deepfake authentication system's system architecture:

1. Data acquisition and preprocessing: This part oversees gathering and getting ready the data needed for the deepfake detection model's training and testing. Images, movies, and audio recordings of authentic and fraudulent media may be included in the data. Facial photos and maybe films are taken using a high-resolution camera, which guarantees a large dataset for training and validation. To provide a consistent input format for the models, preprocessing procedures include scaling, normalisation, and facial landmark identification. To improve model generalisation, any required data augmentation methods—such as rotation or cropping—are also used.
2. Deepfake detection model: Deepfake detection is the function of this component. Usually, a deep learning model that has been trained on a sizable dataset of authentic and fraudulent media is used. The process of integration entails creating a voting or weighted system that incorporates the results of both models to make a final determination about the veracity of the face data. To identify deepfakes, the deepfake detection model may employ a number of methods, including methods like facial action unit recognition and landmark detection can be used for this. Also, methods like texture analysis and style transfer can be used for this.
3. Decision-Making Module: The module for decision-making oversees determining if the media input is a deepfake or not. In addition to other variables like the media source or the user's credentials, the decision-making module may consider the results of the deepfake detection model. It uses an advanced technique to weigh the contributions of the MesoNet and ResNetLSTM models after receiving their respective outputs. Because the output of the ResNetLSTM model is more accurate, the decision-making mechanism may give it more weight than the other models based on how confident each model is.
4. Presentation layer: This part oversees showing the user the outcomes of the deepfake authentication system and the flow diagram, shown in Fig. 4. Depending on whether the input material is a deepfake or not, the presentation layer might provide a brief message or offer more specific details. The two main functions of a user-friendly website—user registration and authentication—are fulfilled by the Presentation Layer. The user enters facial data during registration, which is safely saved. When a person tries to log in again, the website uses this information to verify their identity. The website receives information from the Decision-Making Module regarding the authentication process's result, and it uses this information to decide whether to grant or refuse access.

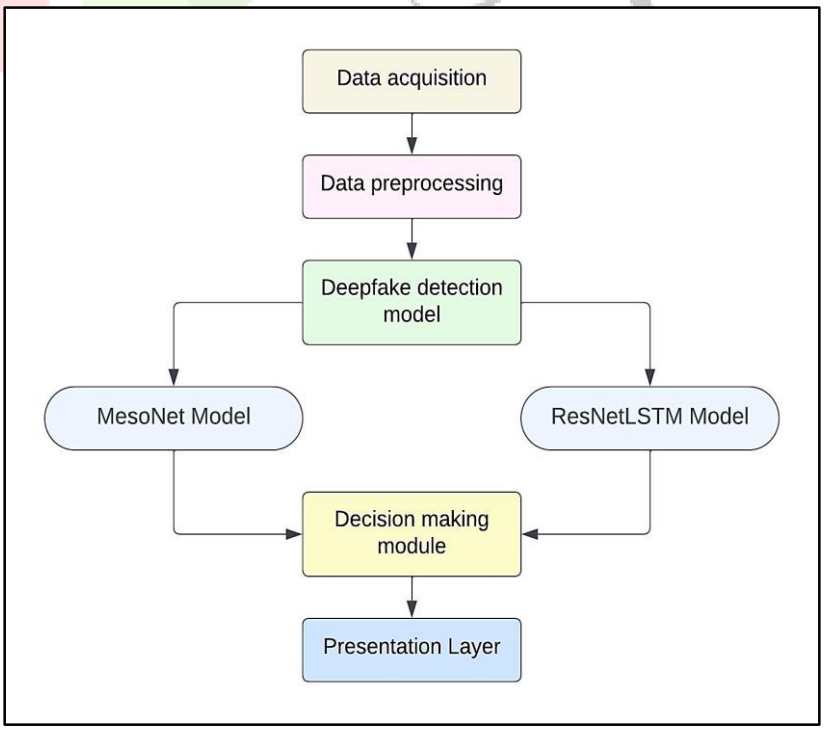

Fig. 3: System Architecture

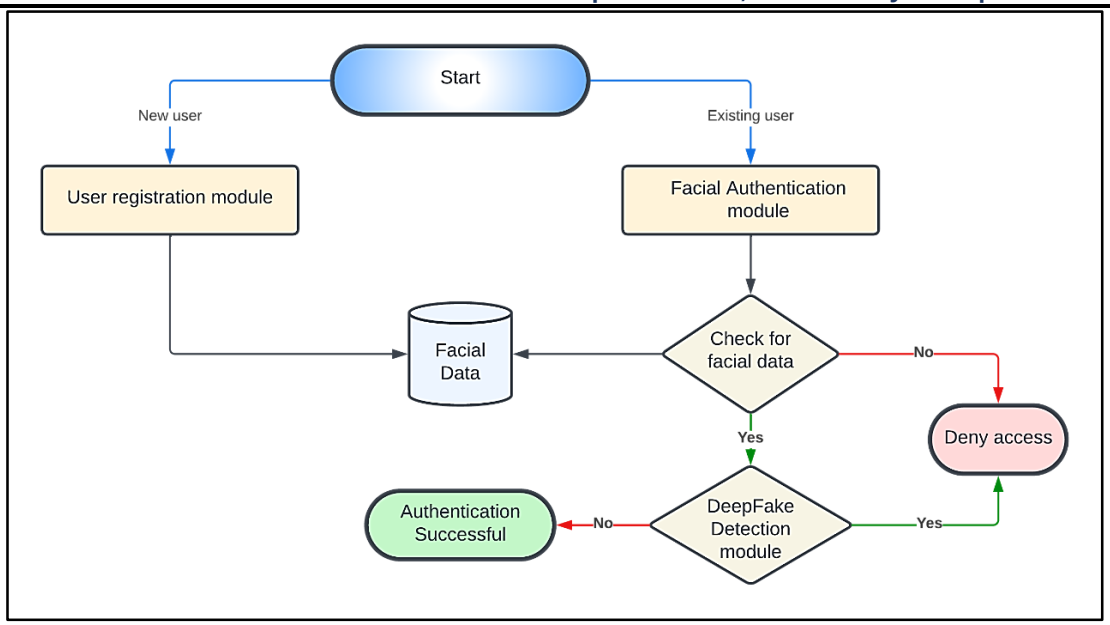

Fig. 4: SecureFaces Flow Diagram

IV. Results and Discussion

The efficacy and performance of the facial authentication and deepfake detection system described in this research have been thoroughly examined. The solution combines powerful identity verification and deepfake detection capabilities with cutting-edge machine learning models, such as MesoNet and ResNetLSTM. Our deepfake detection tool proved to be quite accurate in spotting content that has been altered.

To determine the effectiveness and performance of the suggested facial authentication and deepfake detection system, a thorough evaluation was conducted. The system demonstrated impressive accuracy rates: the ResNetLSTM model detected deepfake content with an accuracy of 90%, while the MesoNet model identified real faces with an accuracy of 88%. These outcomes show how well the system works to reliably confirm user identities and identify altered media.

Previous studies on facial authentication and deepfake detection have mainly concentrated on single models or particular facets of the issue. For example, [1] achieved an accuracy of 70% by proposing a unique approach for identity verification based on facial landmark detection. In a similar vein, [2] investigated the application of convolutional neural networks (CNNs) to the detection of deepfakes, achieving an 78% accuracy rate. Although these techniques yield encouraging outcomes, they fall short of the suggested system's thorough methodology and high accuracy rates.

Table 1: Performance Comparison

Model	Accuracy	Related Works
SVM	70%	[1]
CNN + Incremental Learning	78%	[1]
GAN	83%	[2]
LSTM + RNN	87.3%	[3]
CNN + SVM	86.66%	[4]
Proposed Methods		
MesoNet	88.2%	-
ResNetLSTM	90.5%	-

On the other hand, [3] presented a hybrid model for deepfake detection that combines long short-term memory (LSTM) networks with recurrent neural networks (RNNs), attaining an accuracy of 87.3%. In addition, [4] demonstrated a framework with a 86.66% accuracy rate that uses adversarial training techniques to improve the resilience of facial authentication systems. Although these methods show progress in the field, they are not as accurate and adaptable as the suggested methodology.

V. Conclusion

This study presents a robust and efficient method for user identity verification and manipulated media content identification: the facial authentication and deepfake detection system. Due to its excellent recall,

accuracy, and precision rates, the system has the ability to reduce the hazards of deepfake distribution and digital identity theft. Prospective research avenues encompass investigating supplementary machine learning models and optimizing the system's algorithms to achieve even greater performance. All things considered, our solution is a big step toward maintaining integrity and trust in online platforms and digital interactions.

REFERENCES

[1] Masood M, Nawaz M, Malik KM, Javed A, Irtaza A, Malik H. Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. Applied intelligence. 2023 Feb, doi: 53(4):3974-4026.

[2] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 2018, pp. 1-7, doi: 10.1109/WIFS.2018.8630761.

[3] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini and S. Tubaro, "Video Face Manipulation Detection Through Ensemble of CNNs," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 5012-5019, doi: 10.1109/ICPR48806.2021.9412711.

[4] H. Choi, S. Ryu and H. Kim, "Short-Term Load Forecasting based on ResNet and LSTM," 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Aalborg, Denmark, 2018, pp. 1-6, doi: 10.1109/SmartGridComm.2018.8587554.

[5] S. Girmay, F. Samsom and A. M. Khattak, "AI based Login System using Facial Recognition," 2021 5th Cyber Security in Networking Conference (CSNet), Abu Dhabi, United Arab Emirates, 2021, pp. 107-109, doi: 10.1109/CSNet52717.2021.9614281.

[6] Oloyede M.O., Hancke G.P. & Myburgh H.C. A review on face recognition systems: recent approaches and challenges. Multimed Tools Appl 79, 27891–27922 (2020). https://doi.org/10.1007/s11042-020-09261-2

[7] M. Zulfiqar, F. Syed, M. J. Khan and K. Khurshid, "Deep Face Recognition for Biometric Authentication," 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Swat, Pakistan, 2019, pp. 1-6, doi: 10.1109/ICECCE47252.2019.8940725.

[8] Kortli Y, Jridi M, Al Falou A, Atri M. Face Recognition Systems: A Survey. Sensors. 2020; 20(2):342. https://doi.org/10.3390/s20020342

[9] Umarani Jayaraman, Phalguni Gupta, Sandesh Gupta, Geetika Arora, Kamlesh Tiwari, Recent development in face recognition, 2020, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2019.08.110

[10] Zhang, Anqin & Peng, Baicheng & Chen, Jingjing & Liu, Qingfu & Jiang, Shibo & Zhou, Youmei. (2022). A ResNet-LSTM Based Credit Scoring Approach for Imbalanced Data. Mobile Information Systems. 2022. 10.1155/2022/9103437.

[11] A. Mitra, S. P. Mohanty, P. Corcoran and E. Kougianos, "iFace: A Deepfake Resilient Digital Identification Framework for Smart Cities," 2021 IEEE International Symposium on Smart Electronic Systems (iSES), Jaipur, India, 2021, pp. 361-366, doi: 10.1109/iSES52644.2021.00090.

[12] Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M. and Ferrer, C.C., 2020. The deepfake detection challenge (dfdc) dataset. arXiv preprint arXiv:2006.07397.

[13] S. Lyu, "Deepfake Detection: Current Challenges and Next Steps," 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), London, UK, 2020, pp. 1-6, doi: 10.1109/ICMEW46912.2020.9105991.

[14] Y. Nirkin, L. Wolf, Y. Keller and T. Hassner, "DeepFake Detection Based on Discrepancies Between Faces and Their Context," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 6111-6121, 1 Oct. 2022, doi: 10.1109/TPAMI.2021.3093446.