



Artificial Intelligence in Forensic Science

Nazneen Mansoor¹ and Alexander Iliev^{1,2}✉

¹ SRH University of Applied Sciences, Berlin, Germany

² Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

nazneenm1990@gmail.com

ailiev@berkeley.edu

Abstract. Artificial intelligence is a rapidly evolving technology that is being used in a variety of industries. This report provides an overview of some artificial intelligence applications used in forensic science. It also describes recent research in various fields of forensics, and we implemented a model for a use case in digital forensics.

Keywords: Forensics, CNN, Transfer Learning, Deep Learning, ResNet50, Deepfake detection, AI.

1 Introduction

Forensic science is the study of evidence or the use of scientific methods to investigate crimes. Forensic investigation involves extensive research that includes gathering evidence from various sources and combining it to reach logical conclusions. Data extraction from mysterious sources can be productive and compulsive, but dealing with massive amounts of data can be quite confusing. There has been a wide range of areas from which data has been produced, particularly in the case of forensics, from DNA and fingerprint analysis to anthropology.

With the expanding growth of Artificial intelligence (AI) in all industries, scientists have done various pieces of research to understand how forensic science could benefit from the technologies in the field of AI. Artificial intelligence plays a significant role in forensics because it allows forensic investigators to automate their strategies and identify insights and information. The use of artificial intelligence technology improves the chances of detecting and investigating crimes. Artificial intelligence could assist forensic experts in effectively handling data and providing a systematic approach at various levels of the investigation. This could save forensic researchers a significant amount of time, giving them more time to work on other projects. AI can discover new things by integrating all of the unstructured data collected by investigators.

Unlike traditional forensic identification, the use cases of Artificial intelligence are numerous in the field of forensic science. This report describes a list of Artificial intelligence applications which are used in forensic science, some of the recent studies conducted in this field, and the implementation of the model for an AI use case in digital forensics.

2 Related Works

The drowning diagnosis in forensics is one of the most challenging tasks as the findings from the post-mortem image diagnosis is uncertain. Various studies have been focused on this area of research to address this issue. The researchers in [1] [2] have used a deep learning approach to classify the subjects into drowning or non-drowning cases by using post-mortem lung CT images.

In [1], they used a computer-aided diagnosis (CAD) system which consists of a deep CNN (DCNN) model. DCNN is trained using a transfer learning technique based on AlexNet architecture. For the experiment, they used CT images of post-mortem lungs of around 280 cases in total which included 140 non-drowning cases (3863 images) and 140 drowning cases (3784 images) [1]. They could achieve better results in detecting drowning cases. In [2], the authors used the same dataset as in [1] and used a different deep learning approach based on the VGG16 transfer learning model. VGG16 model [2] showed better performance compared to AlexNet used in [1] for drowning detection.

The automatic age estimation of human remains or living individuals is a vital research field in forensics. Age estimation using dental X-ray images [3] or MRI data [4] is commonly used in forensic identification, but the conventional methods do not yield good results. To enhance the performance, the researchers in [3] implemented a deep learning technique to evaluate the age from teeth X-ray images. They have used a dental dataset that consists of 27,957 labeled orthopantomogram subjects (16,383 for females and 11,574 for males). The accuracy of the age prediction of these subjects is verified using their ID card details. The authors have focused on different neural network elements that are useful for age estimation. This paper is relevant in forensic science for estimating age from panoramic radiograph images. In [4], the researchers proposed an automatic age estimation from multi-factorial MRI data of clavicles, hands, and teeth. The authors in [4] used a deep CNN model to predict age and the dataset consists of 322 subjects with the age range between 13 and 25 years.

Gender estimation is another significant aspect of forensic identification, particularly in mass disaster situations. In [5], the authors proposed a model to estimate the accuracy of gender prediction from Cone Beam Computed Tomography (CBCT) scans. They used linear measurements of the maxillary sinus from CBCT scans and principal component analysis (PCA) was done to lower the dimensionality.

With the development in the field of Artificial intelligence, face recognition is a significant research topic that can be beneficial in digital forensics. The researchers in [6] have implemented a deep learning system that can analyze image and video files retrieved from forensic evidence. The proposed model in [6] could detect faces or objects from the given images or videos. The model is trained using YOLOv5 algorithms which is a novel convolutional neural network (CNN) that can be used to detect faces or objects for real-time applications. From their study, it's proven that algorithms trained using deep learning techniques are highly preferred and useful for forensic investigations.

Most of the data extracted from digital forensics contain unstructured data like photos, text, and videos. Data extracted from text play a significant role in digital forensics. Natural language processing (NLP) is an interesting research topic in the field of AI which can be used to provide relevant information in forensics too. In [7], the

researchers developed a pipeline to retrieve information from texts using NLP and built models like named entity recognition (NER), and relation extraction (RE) in any language. From their experimental results, it's proven that this solution enhances the performance of digital investigation applications. In [8] and [9], the researchers used CNN models to identify the Deepfake media in digital forensics.

3 Applications Of AI In Forensic Science

Artificial intelligence assists forensic scientists by providing proper judgment and techniques to have better results in various areas. It involves forensic anthropology by finding out the skeletal age or sex, dealing with a large amount of forensic data, associating specific components from the images, or discovering similarities in place, communication, and time. Some of the AI use cases in the field of forensic science are listed below.

3.1 Pattern Recognition

Pattern recognition is one of the main applications of AI in forensic science which deals with identifying certain types of patterns within a massive amount of data [11]. It can include any pattern recognition such as images of a person, or place, forming sequences from a text like an email, or messages, and other audio patterns from sound files [11] [12]. This pattern matching is based on solid evidence, statistics, and probabilistic thinking. AI helps in providing better ideas for identifying the trends with complex data accurately and efficiently. It also helps the detectives to find the suspect by providing information about past criminal records. The methodology which was used for face image pattern recognition is shown in Figure 1.

3.2 Data Analysis

Digital forensics is a developing field that requires complex and large dataset computation and analysis [11]. In digital forensics, scientists manage to collect digital shreds of evidence from various networks or computers [14]. These pieces of evidence are useful in various areas of investigation. Artificial intelligence acts as an efficient tool in dealing with these large datasets [14]. With the help of AI, a meta-analysis of the data extracted from different sources can be conducted. This meta-data can be transformed into an understandable and simplified format within a short time [11].

3.3 Knowledge Discovery

Knowledge discovery and data mining are the other areas in which artificial intelligence is used. Knowledge discovery is the technique of deriving useful information and insights from data. Data mining includes AI, probabilistic methods, and statistical

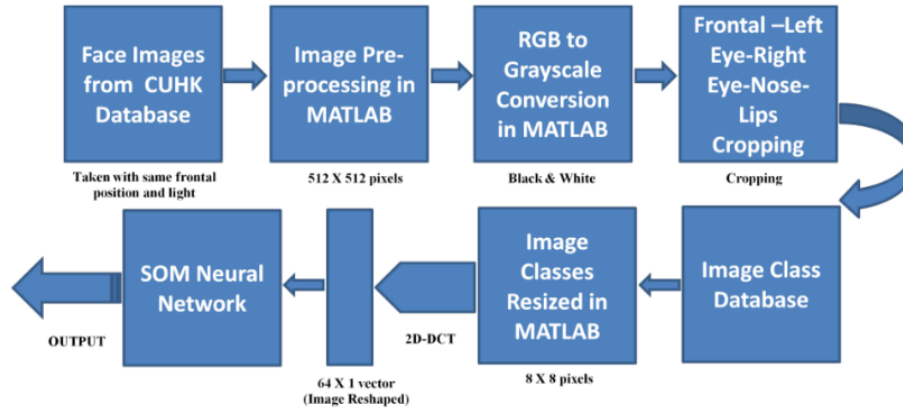


Fig. 1. Methodology for forensic face sketch recognition [17]

analysis, which can gather and analyze large data samples. Forensic scientists could use AI to investigate various crimes using this approach to discover patterns.

3.4 Statistical Evidence

In forensic science, strong statistical evidence is required to support the arguments and narration [13]. AI helps in building graphical models that can be used to support or disprove certain arguments, thereby making better decisions. There are various computational and mathematical tools in AI that helps to build significant and statistically relevant evidence.

3.5 Providing Legal Solutions

The scientific methods provided by forensic statistics support the legal system with the necessary evidence [13]. With more comprehensive and sophisticated information databases, artificial intelligence aids the legal community with better solutions.

3.6 Creating Repositories

With the increasing demand in the case of storage capacity, forensic investigators find it difficult to store and analyze data related to forensic science [13]. The storage issues can be rectified by building online repositories using AI. These repositories are useful to store digital forensic data, properties, investigations, and results [14].

3.7 Enhance Communication Between Forensic Team Members

In a forensic investigation, it's necessary to maintain strong communication between forensic statisticians, criminal investigators, lawyers, and others [13].



9

Miscommunication between these teams could result in misinterpretation of data which leads to wrong decisions or unfair justice. Artificial intelligence assists to bridge the communication gap between various teams in the forensic field.

4 Proposed Methodology

We have tried to implement a system to detect Deepfake images which can be widely used in digital forensics. With the advancement of technology and the ease of creating fake content, media manipulation has become widespread in recent years [10]. These fake media contents are generated using advanced AI technologies like deep learning algorithms and hence they are known as “Deepfakes” [16]. Therefore, it has become difficult to differentiate the original media from the fake ones with our naked eye. There are numerous Deepfake videos or images which are circulated across social media and these manipulated videos or images can be a threat to society and could lead to an increase in the number of cybercrimes. With the increasing number of Deepfake content, there is a high demand for developing a system that could detect Deepfake videos or images. Artificial intelligence contributes efficiently to detecting these manipulated multimedia content [16].

The proposed system uses a neural network with a transfer learning technique that is based on ResNet50 architecture to train our model.

4.1 Data Collection and Pre-processing

The dataset selected to train the model is the Celeb-DF dataset. It includes 590 celebrity videos and 300 additional videos that are downloaded from YouTube. Also, it contains 5639 synthesized videos generated from celeb real videos. The videos are converted into frames using the cv2 python library. From the frames, faces are cropped and in total, there were 19,457 images. For training the model, we have used 80% of the data which is around 15565 images, 10% for validation, and 10% for testing purposes.

Fig. 2. Fake Images [15]



Fig. 3. Real Images [15]

Fig 2 and Fig 3 denote an example of a few of the fake and real images which are part of our dataset. Pre-processing of the images is done using the Image data generator class of Keras applications.

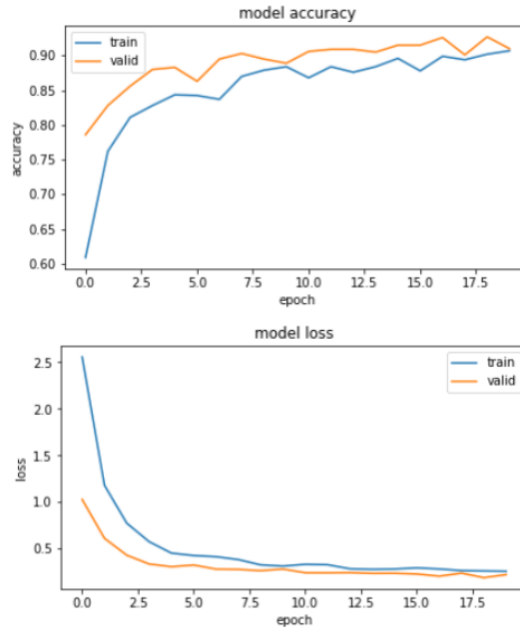
4.2 Training The Model

The model is developed using ResNet50 architecture from the Python Keras application and there are two additional dense layers used in our model with a dropout of 50%. For activation functions, rectified linear unit (ReLU) is used in the hidden layer and the sigmoid function is used in the output layer. The number of epochs for training the models is set to 60 respectively. For compiling the models, we have opted for Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.9 and a learning rate of 0.001 and used binary cross-entropy as the loss function

5 Results

Accuracy is the performance metric that is used to evaluate the efficiency of our model. There are two graphs attained that show the accuracy and loss during training the

Fig. 4. Training and Validation learning curves (accuracy and loss)



model. Fig 4 illustrates the training and validation curves with the accuracy and loss of the trained model. The confusion matrix of the model is depicted in Fig 5 which shows the actual and predicted number of fake and real images. From the confusion matrix,

the test accuracy is evaluated as 92%. In Fig 6, the example of a predicted class of an image using the model is depicted.

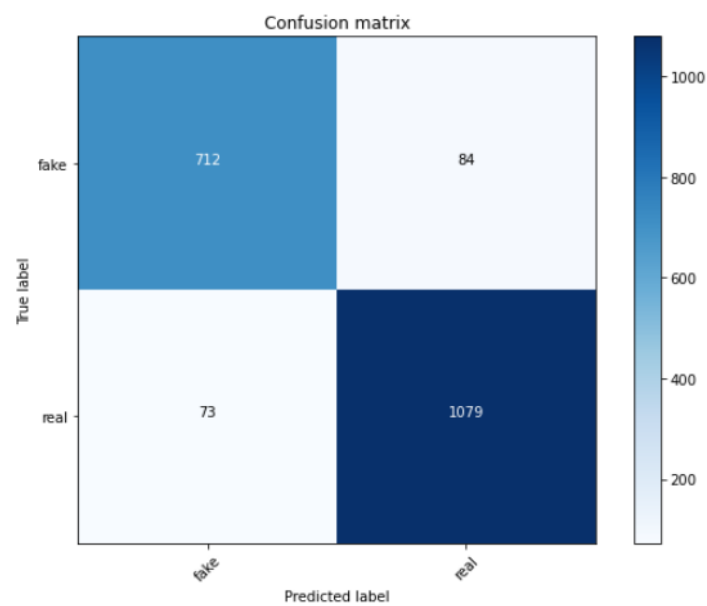


Fig. 5. Confusion matrix of the model

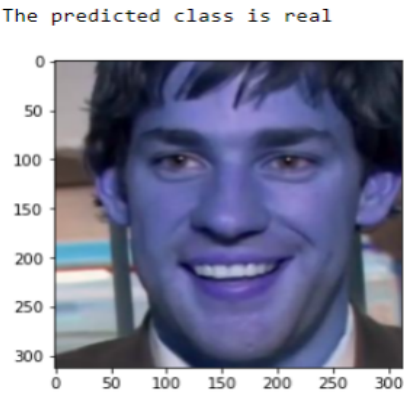


Fig. 6. Example of a predicted image

6 Conclusion

Artificial intelligence is rapidly becoming the most important applied science in all fields. This report summarizes a few of the various AI applications in forensics, and it could be concluded that AI can help forensic experts or investigators reduce the time taken on various tasks and thereby improving their performance. Some of the AI applications that are prevalent in the field of forensic science are pattern recognition, handling large amounts of data, and providing legal solutions. Researchers have made various studies to understand how forensic science benefits from AI technologies for different use cases such as drowning diagnosis from images, age, and gender estimation. As part of our study, we have tried to implement a neural network model based on ResNet50 architecture to detect Deepfake images and it could be useful in digital forensics. The proposed model can differentiate real and fake images with an accuracy rate of around 92%.

References

1. N. Homma et al., "A Deep Learning Aided Drowning Diagnosis for Forensic Investigations using Post-Mortem Lung CT Images," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 1262-1265, doi: 10.1109/EMBC44109.2020.9175731
2. A. H. Qureshi et al., "Deep CNN-Based Computer-Aided Diagnosis for Drowning Detection using Post-mortem Lungs CT Images," 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021, pp. 2309-2313, doi: 10.1109/BIBM52615.2021.9669644.
3. W. Hou et al., "Exploring Effective DNN Models for Forensic Age Estimation based on Panoramic Radiograph Images," 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1-8, doi: 10.1109/IJCNN52387.2021.9533672.
4. D. Štern, C. Payer, N. Giuliani and M. Urschler, "Automatic Age Estimation and Majority Age Classification From Multi-Factorial MRI Data," in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1392-1403, July 2019, doi: 10.1109/JBHI.2018.2869606.
5. A. Al-Amodi, I. Kamel, N. H. Al-Rawi, A. Uthman and S. Shetty, "Accuracy of Linear Measurements of Maxillary Sinus Dimensions in Gender Identification Using Machine Learning," 2021 14th International Conference on Developments in eSystems Engineering (DeSE), 2021, pp. 407-412, doi: 10.1109/DeSE54285.2021.9719421.
6. S. Karakuş, M. Kaya, S. A. Tuncer, M. T. Bahşi and M. Açıkoğlu, "A Deep Learning Based Fast Face Detection and Recognition Algorithm for Forensic Analysis," 2022 10th International Symposium on Digital Forensics and Security (ISDFS), 2022, pp. 1-6, doi: 10.1109/ISDFS55398.2022.9800785.
7. F. B. Rodrigues, W. F. Giazza, R. de Oliveira Albuquerque and L. J. García Villalba, "Natural Language Processing Applied to Forensics Information Extraction With Transformers and Graph Visualization," in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2022.3159677.

8. Jadhav, Ekta & Singh Sankhla, Mahipal & Kumar, Rajeev. (2020). Artificial Intelligence: Advancing Automation in Forensic Science & Criminal Investigation. Seybold Report. 15. 2064-2075.
9. Vurimi Veera Venkata Naga Sai Vamsi, Sukanya S. Shet, Sodum Sai Mohan Reddy, Sharon S. Rose, Sona R. Shetty, S. Sathvika, Supriya M. S., Sahana P. Shankar, Deepfake detection in digital media forensics, Global Transitions Proceedings, Volume 3, Issue 1, 2022, Pages 74-79, ISSN 2666-285X, <https://doi.org/10.1016/j.gltp.2022.04.017>.
10. M. T. Jafar, M. Ababneh, M. Al-Zoube and A. Elhassan, "Forensics and Analysis of Deepfake Videos," 2020 11th International Conference on Information and Communication Systems (ICICS), 2020, pp. 053-058, doi: 10.1109/ICICS49469.2020.239493.
11. NEWS MEDICAL LIFE SCIENCES, "AI in Forensic Science" <https://www.news-medical.net/life-sciences/AI-in-Forensic-Science.aspx> (last accessed 2022/06/30)
12. Jadhav, Ekta & Singh Sankhla, Mahipal & Kumar, Rajeev. (2020). Artificial Intelligence: Advancing Automation in Forensic Science & Criminal Investigation. Seybold Report. 15. 2064-2075.
13. Mohsin, Kamshad. (2021). Artificial Intelligence in Forensic Science. SSRN Electronic Journal. 10.2139/ssrn.3910244.
14. Siddhant Gupta (2020). ARTIFICIAL INTELLIGENCE IN FORENSIC SCIENCE. IRJET. e-ISSN: 2395-0056
15. @inproceedings{Celeb_DF_cvpr20, author = {Yuezun Li, Xin Yang, Pu Sun, Honggang Qi and Siwei Lyu}, title = {Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics}, booktitle= {IEEE Conference on Computer Vision and Pattern Recognition (CVPR)}, year = {2020} } - celeb df (last accessed 2022/07/05)
16. Karandikar, Aarti. (2020). Deepfake Video Detection Using Convolutional Neural Network. International Journal of Advanced Trends in Computer Science and Engineering. 9. 1311-1315. 10.30534/ijatcse/2020/62922020.
17. @inproceedings{Srivastava2013ForensicFS, title={Forensic Face Sketch Recognition Using Computer Vision}, author={Vineet K. Srivastava}, year={2013}}