# Trustworthy AI in Biometrics: A Comprehensive Review of Trends and Challenges

Shefali Arora[1,*], Kanu Goel[2], Shikha Gupta[3], Ruchi Mittal[4] and Avinash Kumar Shrivastava[5]

[1]*Department of Computer Science and Engineering, Dr BR Ambedkar National Institute of Technology, Jalandhar, India*
[2]*Department of Computer Science and Engineering, Punjab Engineering College, Chandgarh, India*
[3]*Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India*
[4]*Iconic Data, Japan*
[5]*Dept. of Quantitative Techniques & OM, International Management Institute, Kolkata, India*
*E-mail: chouhansa@nitj.ac.in*
[*]*Corresponding Author*

Received 28 August 2024; Accepted 23 December 2024

## Abstract

Deep Learning (ML) and Artificial Intelligence (AI) are rapidly spreading over many application domains. However, the creation of intelligent systems is constrained by inherent flaws in the learning algorithms that are employed. One major barrier to the application of these techniques is the unpredictable nature of model performance. The reliability of a model is determined by its capacity to remove biases, elucidate findings, and adapt to changes in input data. The idea of trustworthy AI uses a variety of machine learning techniques to explain a model's decision-making process, hence boosting

*Journal of Graphic Era University, Vol. 13_1,* 91–118.
doi: 10.13052/jgeu0975-1416.1315

user confidence in the model's output. The significance of reliable AI in the context of biometrics is the main topic of this study. Our survey identifies the several types of reliable AI that can be used to increase the dependability of the choices made.

**Keywords:** Trustworthiness, bias, explainability, biometrics, security, performance.

## 1 Introduction

The Oxford English Dictionary and the Cambridge Dictionary define "trustworthy" as something that is dependable and deserving of confidence. According to the Oxford English Dictionary, trust is a strong belief in the dependability, truthfulness, or competence of an individual or entity, but the Cambridge Dictionary describes it as confidence in the reliability of someone or something. Generally speaking, trust is a widely accepted idea in human society and serves as a vital pillar for the ongoing development of civilization. There will always be external hazards in our surroundings since humans and other entities cannot be completely controlled.

Nonetheless, we deliberately expose ourselves to potential hazards in order to preserve connections because we have faith in these parties. Trust is especially important in relationships because it allows people to coexist peacefully and productively, and it is the foundation of successful partnerships.

Ensuring that AI systems are impartial is essential to developing a reliable system. By introducing biases into their outputs, a number of AI systems, ranging from language translation to facial recognition, have shown their shortcomings [15]. The inability of many of these systems to understand and justify their decision-making procedures jeopardizes the fairness of their results. Even in cases where the AI system demonstrates exceptional accuracy and efficiency, this drawback persists. On the other hand, black box AI system decisions are vital in critical scenarios like video monitoring and medical diagnostics.

Such systems allow for adversarial attacks and compromises of authentic user identities. Building strong frameworks for the creation of reliable AI systems is therefore essential. There is much room for development as biometric authentication is employed increasingly frequently in a range of applications utilizing different human characteristics. Recently, deep learning and artificial

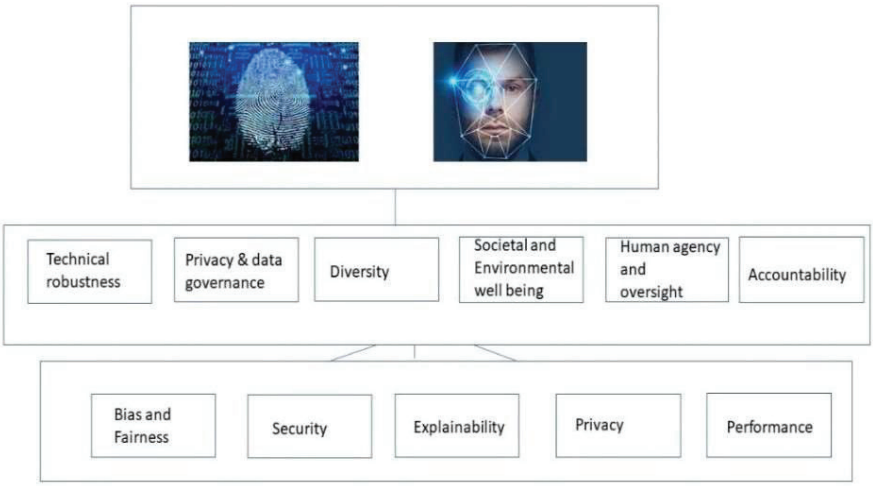


**Figure 1** Several dimensions in trustworthy AI systems.

intelligence techniques have showed promise in a number of fields, including biometrics. However, striking a balance between explainability and accuracy is challenging due to the complexity of biometric systems, which are often represented by complicated neural network models like Convolutional Neural Networks (CNNs).

The growing use of facial recognition software, which is common but vulnerable to security lapses, serves as a reminder of this issue. It is quite challenging to explain models such as VGG-19, a convolutional neural network with 19 layers and more than 140 million parameters, due to their tremendous complexity. Understanding the reasons behind errors is made easier by incorporating reliability into biometric systems. The topic of trustworthy AI in biometrics has already been covered in existing literature, therefore a thorough review of achievements, a determination of research gaps, and a delineation of future paths are all necessary. Figure 1 shows the several aspects of reliable AI in biometrics. These illustrate the different facets that fall under the purview of explainable AI, including performance, privacy, bias and fairness, and characteristics such as privacy, governance, accountability etc.

The paper summarizes work done in trustworthy AI with identification of various human traits such as fingerprint, face, iris etc. We analyze the existing research done to review the issues and emerging trends.

### 1.1 Research Gaps and Objectives

Understanding the function of reliable AI in biometrics is the primary goal of this review. The authentication procedure of biometric systems should be secure and accurate. Further investigation is also required into the ethical issues surrounding the combination of these two technologies [17]. There are a number of problems, including privacy difficulties and the possibility of bias in AI algorithms that make judgments for biometric systems. Malicious attempts by attackers to manipulate such systems through carefully designed inputs are also a possibility. All of the topics related to biases, security and privacy, and the rationale behind the system's decisions are covered by trustworthy AI [7]. There are certain research gaps which have been kept in mind while framing the research questions:

(1) Despite the widespread usage of biometric systems, little research has been done on the ethical concerns associated with integrating trustworthy AI with biometric frameworks, particularly with regard to responsible data collection and use.
(2) Many AI-powered biometric systems are susceptible to computational biases that lead to skewed outcomes. More research is needed to develop and assess algorithms that lessen biases, especially those that affect underrepresented groups.
(3) It is difficult to comprehend and defend the decisions made due to the opaque AI algorithms used in biometric systems. To find out how trustworthy AI can make these systems simpler to comprehend and analyze, more research is needed.
(4) Adversarial input attacks and spoofing are threats to AI-driven biometric systems. Research is needed to provide more security protocols.

In this paper, we comprehensively review the various AI algorithms which have been integrated with biometric frameworks in order to deal with one of the aforesaid dimensions of trustworthiness. We highlight the existing works in terms of the technique used, the biometric trait focused on, the strengths of using the proposed framework. We also discuss the findings in terms of the ongoing trends in this domain, also analyzing the challenges and future directions that can be taken up by researchers. Moreover, the inclusion of trustworthiness should be guided by ethical guidelines that concern issues related to data collection, mitigation of bias and responsible usage of biometric data.

By addressing these research gaps and objectives, the comprehensive review can contribute to the advancement of trustworthy AI in biometric

systems, ensuring their accuracy, security, and ethical use in various applications. Neural networks are the base of all the AI technologies which are integrated with biometric systems to add trustworthiness. Several ethical concerns may arise in various situations such as:

1. **Facial Recognition in Law Enforcement:** The application of AI-powered facial recognition technology has sparked ethical questions. For example, algorithmic biases have resulted in false arrests, which have disproportionately affected minority communities. This highlights the need for moral AI to avoid biased results in biometric systems intended for public safety.
2. **AI and Biometric Data Privacy:** There have been instances where face recognition photos were taken from a third-party contractor as a result of a hacking. This hack brings to light moral questions about the security of biometric data and emphasizes the need for more reliable, strong AI systems to safeguard private data.
3. **AI for Health Biometrics:** AI-driven biometric technologies are being utilized more and more in the medical field for patient tracking and identification. On the other hand, moral questions about informed consent and data security surface. The use of AI in COVID-19 thermal scanning and facial recognition apps, for example, has prompted concerns about data usage and privacy in some countries. This highlights the significance of ethical frameworks when integrating AI with biometric systems in health.

Figure 2 depicts the structure of a neural network which is used to perform various tasks related to explainability in biometrics.

The remaining work is organised as follows. Section 2 discusses the dimensions of trustworthy AI and deep learning technology. Section 3 describes the review technique and responds to the research issues posed in this article by comparing existing literature work throughout time. Section 4 discusses the entire literature review's findings, conclusions, and recommendations for further work. Section 5 concludes the research.

## 1.2 Research Questions

The formulated research questions are as follows:

RQ1. What are the various methods and frameworks available in the existing literature to add trustworthiness in such systems?
RQ2. What are the various frameworks available in trustworthy AI?

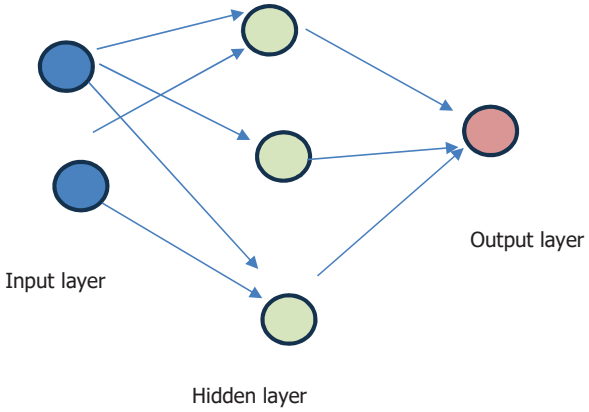


**Figure 2** Structure of a neural network.

RQ3. What are the various advantages and challenges of adding trustworthiness to biometric systems?

## 2 Background of Trustworthy AI in Biometrics

As per the European AI act [35] it is necessary to adhere to these seven guidelines to ensure trustworthiness in biometric systems: (1) Human agency and oversight, including fundamental rights) Resilience to attacks and technical robustness (3) Respect for privacy and integrity of data (4) Explainability and transparency (5) Fairness, diversity, and no discrimination (6) Societal and environmental well-being. The several dimensions which add trustworthiness to a biometric system are categorized in the paper as follows:

- Bias and Fairness – In an automated biometric system, three potential sources of bias exist: the data, the algorithm, and the assessment technique. Data bias manifests itself at numerous stages of data development, including data collection, processing, and annotation. Bias can be classified into different categories from distinct perspectives. Indirect bias refers to the phenomena in which non-sensitive and neutral traits create an unsatisfactory outcome. Explicit or direct bias occurs when an attribute openly generates an undesirable consequence.
- Performance – There is a need to integrate two-track strategy in AI systems such as biometric authentication systems. One track would concentrate on the certification of AI applications through a conformity assessment agency utilising objective criteria. There is a need to build

a new track for AI engaging with humans and also increasing the performance of such systens. The role of process mining is useful for increasing the performance and trustworthiness of AI systems such as biometric authentication systems is discussed in a study of obstacles and opportunities in the domain of trustworthy AI.

- Security and Privacy – In cyber security applications, most models are called black boxes. For example, introducing explainability to biometric systems helps to increase security. Several recent research papers propose ways for generating reasons for faulty intrusion detection system applications. Several more research papers provide effective techniques to improving client security while using biometric authentication in a variety of applications. The usage of biometric technologies is intended to prevent illegitimate use. However, the biometric templates may not be secure. This could result in a plethora of security vulnerabilities, including denial-of-service attacks and client privacy violations.
- Explainability – To improve the trustworthiness of biometric systems, researchers have utilised a variety of approaches to improve explainability and interpretability. Explainability can be divided into three categories:
  (a) Intrinsic model This category includes model-specific and intrinsic explanations. The same interpretability strategies cannot be used by different classifiers.
  (b) Model-independent explanations: These methods are concerned with black-box well-trained AI models. These techniques are used to interpret previously trained models. As a result, it is sometimes referred to as the black box explainability method. Local Interpretable Model-Agnostic Explanations (LIME) is one of the most notable studies in this category. A deep learning model processes the obtained data, identifying patterns that can be applied to a variety of use cases. The following section provides an overview of deep learning models that have formed patterns and made judgements using biometric system data. Figure 3 depicts the process of implementing deep learning models in biometric systems. This pipeline involves training followed by validation and evaluation.

The various deep learning frameworks which are employed to add trustworthiness to biometric systems are:

- Convolutional Neural Networks – CNNs are a type of neural networks that includes both convolutional and fully connected layers. Their

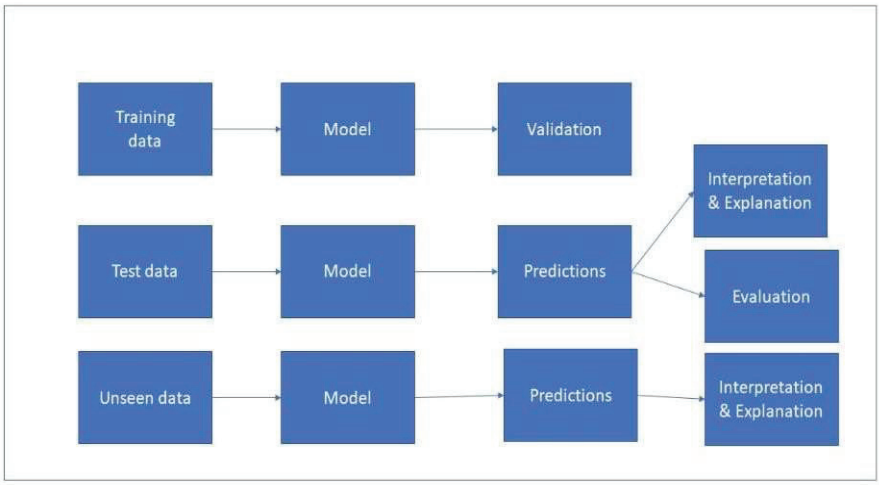


**Figure 3** Trustworthy AI in biometric systems.

weights influence how neurons communicate with one another. Each neuron in the network calculates the dot product of the input x, weight w, and bias b to determine the output. g is an activation function that aids in adding non-linearity and lowering the number of parameters. It can be sigmoid or the Rectified Linear Unit.

- Deep Belief Networks – The deep belief network (DBN) is a type of deep neural network that consists of numerous layers of models with both directed and undirected edges. In a deep belief network, there are several hidden units. Units are not related to one another, but layers are.
- Long-Short Term Memory Networks – LSTMs are a particular variety of RNNs with three gates: input gates, output gates, and forget gates. In LSTM model, three things must be decided i.e. how added information should be remembered and how much previous information should be overlooked.

## 3 Methodology

The protocol used for the purpose of review is depicted in Figure 4. The review focuses on English-language papers from Web of Science journals and excludes irrelevant articles. Each review articles address atleast one research question.

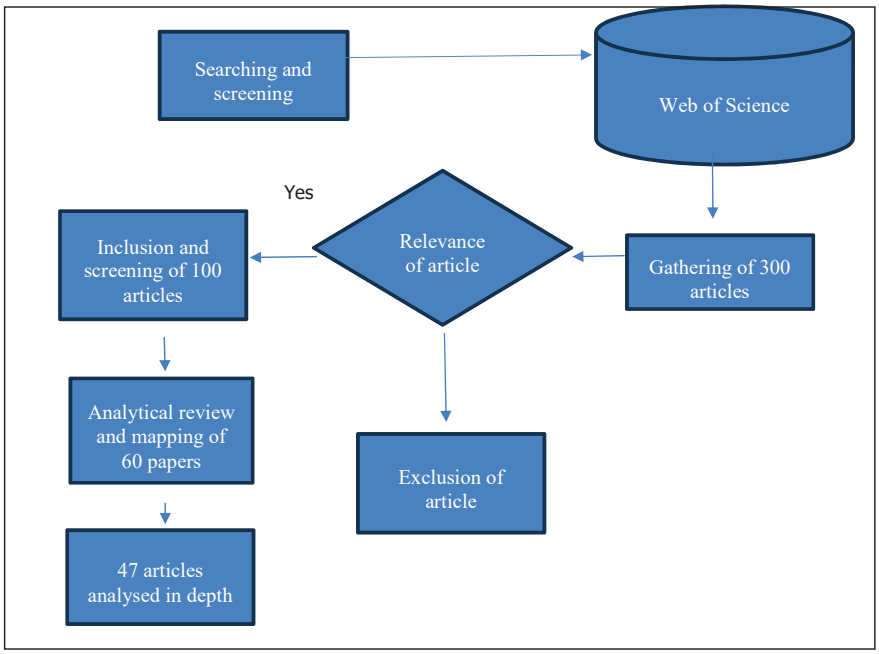


**Figure 4** Detailed methodology of review.

The query strings were merged to search multiple databases for relevant publications about trustworthy AI, biometrics, deep learning, and applications. Out of 300 papers obtained, 100 were used for the preliminary review. The chosen literature work's eligibility was determined using the framed review process. The primary focus was on the characteristics of trustworthy AI mixed with deep learning to answer a specific research issue.

In Section 3, we analyze 47 papers that discuss the research done in the domain of trustworthy AI in biometrics.

### 3.1 RQ1: Current Research Trends

Table 1 lists the various existing works which highlight the research trends in the domain of explainable AI in biometrics. The table also depicts their strengths and the explainability aspect used in the work.

The various research trends in the domain of trustworthy AI in biometrics have been depicted in Figure 5. A few research papers which deal with

**Table 1** Existing research work in the domain of Trustworthy AI in biometrics

| Work | Framework Used | Strengths | Dataset | Biometric Trait | Trustworthy Factor |
|---|---|---|---|---|---|
| [5] | Analysis of edge-AI in human centric machine learning applications with explanations generated using SHAP that perform emotion recognition | The input quantization is significantly biassed against dark skin faces, and the observed patterns may be due to low-contrast features of dark skin faces. | AffectNet dataset | Face | Bias |
| [34] | Novel learning scheme which extracts features from neural models, preserves identities and removes demographic knowledge | Prevents potential biases from affecting decision making in various ethnic groups. | Balanced Faces in The Wild | Face | Bias |
| [26] | Evaluation of bias in face recognition systems based on the level obtained for trained model | Analysis of patterns that suggest bias in recognition of people from different ethnic backgrounds and age groups | CMU Multi-PIE dataset, Craniofacial Longitudinal Morphological (MORPH) Album-2 dataset, Racial Faces in the Wild (RFW) | Face | Bias |
| [11] | Analysis of four metrics in datasets: Sustainability, Accuracy, Fairness, Explainability | Dataset used is biased towards men as compared to women | AML Banking dataset | Soft biometrics (Gender) | Bias, Performance |
| [40] | Estimation of pixel quality of facial images using a quality regression model to understand image quality and generate interpretations | Increases interpretability of face images | Adience, VGGFace2 | Face | Explainability |

| | | | | | |
|---|---|---|---|---|---|
| [23] | Use of homomorphic encryption for security of different traits | Combines the benefits of biometrics and cryptography | ORL and FVC2002 | Face, Fingerprint | Security& Privacy |
| [18] | A comprehensive survey of approaches used for face liveness detection | Review of various explainable frameworks | OULU-NPU, Replay-Attack, MSU MFSD etc. | Face | Explainability, Performance |
| [36] | Use of ensemble face recognition approaches based on three-way decision making | Explainable authentication with superior accuracy | Celebfaces | Face | Explainability, Performance |
| [31] | Use of LIME with pre-trained CNN models to make visual explanation maps | Interpretation of visually significant features using LIME | Yale, AT &T dataset, and Labeled Faces in the Wild (LFW) | Face | Explainability |
| [20] | Use of conditional autoencoders to detect aging in individuals | Interpretation of face areas using LIME on a pre-trained age classifier | CACD, UTKFace | Face | Explainability |
| [35] | Use of XQNets for obtaining quality value along with other environmental attributes for facial detection | Fast and explainable approach for face image quality assessment | LFW and SCFace | Face | Explainability |
| [29] | Use of LSTM and Grad-CAM with modified gate function to detect presentation attacks | Improved accuracy, security and usable explanations | CASIA-FA, Replay Attack, MSU-MFSD, and HKBU | Face | Explainability, Secu-rity&Privacy, Performance |
| [39] | Use of sensitive loss function based on triplet loss and triplet generator. Use of Discrimmation aware deep learning-based face recognition using pre-trained CNN models | Improved accuracy and fairness in recognizing people from different backgrounds and ethnicities. | Three public databases of 64000 identities from different demographic regions | Face | Bias, Performance |

(*Continued*)

**Table 1** Continued

| Work | Framework Used | Strengths | Dataset | Biometric Trait | Trustworthy Factor |
|---|---|---|---|---|---|
| [43] | Comparison of diverse sclera segmentation techniques to understand biases introduced by demographic and environmental factors | Negative correlation between amount of bias and segmentation performance. | SLD, MASD, SMD and other datasets | Iris | Bias |
| [38] | Monitoring and authentication of individuals entering public stadiums, airports and other public places | Countermeasures to secure open public spaces and semi-open public spaces, along with analysis of ethical and societal risks | Custom dataset | Face | Bias, Security&Privacy |
| [30] | Use of multimodal self-organized neural networks combining EEG and keystroke | High accuracy of 98% | Data collected in ten sessions | EEG | Performance |
| [1] | Optimized AI algorithms (ANFIS) and SVM with error correction codes along with subtractive clustering | High accuracy of 99% and reduction in bias based on demographic cues | Intrinsic hand dataset | Soft biometrics (size, width, height, sex) | Performance, Bias, Security and Privacy |
| [2] | Use of CNN and GradCAM to produce visual explanations for human activity recognition | Improves performance and bias | UniMiB-SHAR | Behavioral biometrics | Explainability, Performance, Bias |
| [25] | Analysis of risks in facial recognition in public spaces | Discussion about enforcement of legal rules and regulations for digital identity systems | Brazilian data framework | Faces | Security & Privacy |

| | | | | | |
|---|---|---|---|---|---|
| [15] | Survey of problems leading to bias in biometric systems, along with other factors such as security, privacy and explainability | Highlights the challenges and issues in these domains | Survey of various datasets | Face, fingerprint, iris | Bias |
| [9] | Analysis of biometric tools that are used in the modern age for identification | Highlights the potential data breaches, loss of data rights and information | | | Security &Privacy |
| [40] | Use of classifiers to analyse face embeddings and encodings for hair color, hair styles, beards and accessories | Works on different soft biometric traits. | LFW and CelebFaces | Face | Bias |
| [28] | Incorporation of Learning From Interpretation transition technique for fair recruitment based on machine learning tool for ranking CVs based on gender and ethnicity | Provides logical framework to justify and explain how gender and ethnicity influence greater salaries. | Data of adult incomes from the US census | Soft biometric traits (gender& ethnicity) | Bias |
| [23] | Use of GradCAM framework to visualize explanations | Explainable Cosine similarity technique to give meaningful explanations. | LFW dataset | Face | Explainability |
| [23] | Use of improved cosine similarity function for explainable authentication metric | Improved face verification results | LFW | Faces | Explainability, Performance |
| [16] | Use of VGG-Net for spoofing detection | Accuracy of 98.3% | FVC 2002, 2004,2006 | Fingerprint | Performance |

(*Continued*)

**Table 1** Continued

| Work | Framework Used | Strengths | Dataset | Biometric Trait | Trustworthy Factor |
|---|---|---|---|---|---|
| [41] | Unsupervised fair score normalization approach for reduction of bias in facial recognition | Reduces demographic biases, for example, by up to 82.7% when gender is taken into account. Furthermore, it mitigates bias more consistently than previous work. | Colorferet | Face | Bias |
| [14] | Two-factor biometric authentication by using fuzzy based procedures for feature extraction | Intelligent security solution with better performance | Custom dataset | Fingerprint | Security& Privacy, Performance |
| [10] | A comprehensive survey of possible biases in face recognition systems | Discussion of technical and social issues | | | Bias |
| [4] | Facial analytics to detect criminality from faces | Consideration of factors which can lead to bias and conclusion of explainable AI picking up pace | NIST | Face | Bias, Explainability, Security&Privacy |
| [44] | Recognition based on pixel-wise annotated features of the schlera | Effective understanding of the schlera structure | SBVPI | Iris | Performance |
| [19] | Use of frequency-based saliency maps generation from face based on entropy, brightness and local binary patterns | Superior classification results in terms of performance and interpretation | Cohn-Kanade database, FG-NET FEED database, and Dartmouth database | Face | Explainability, Performance |

| | | | | | |
|---|---|---|---|---|---|
| [46] | Use of graph fusion based technique to resolve feature space mismatch in fingerprints | High accuracy of 99% | Home-made trimodal dataset | Fingerprint | Performance |
| [47] | Proposed lightweight encryption scheme for protection of biometric templates | No critical information revealed from templates under active and passive attacks | Fingerprint database | Fingerprint | Security& Privacy |
| [48] | Encryption of query data and submission to cloud | Proposed framework is secure against attacks | Database with several traits | Fingerprint, iris, voice and face | Security& Privacy |
| [27] | Use of CNN and SVM by combining local and global iris regions to detect presentation attacks | Improved performance and security against attacks | Warsaw-2017 | Iris | Security& Privacy, Performance |
| [24] | Use of ensemble biometric authentication using secret key and | Analysis of False Accept and False Reject probabilities | Public Database | | Security& Privacy, Performance |

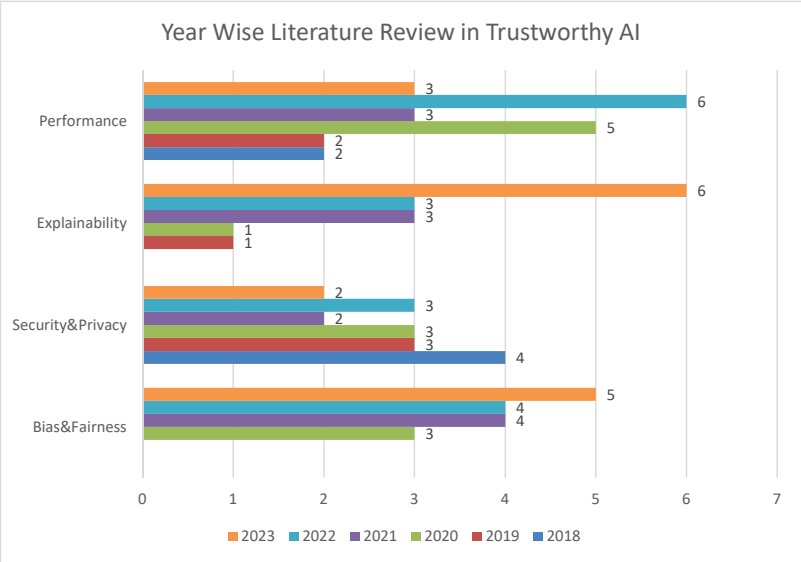


**Figure 5** Year-wise analysis of published work.

various dimensions of trustworthy AI have been counted separately in this figure.

### 3.2 RQ2: Existing Frameworks in Trustworthy AI

These days, several tools and frameworks help us understand and interpret the predictions made by AI models. Using these tools, we can debug the model's performance and understand the decisions in a better manner. The various popular tools are as follows:

(1) LIME (Local Interpretable Model-Agnostic Explanations): IME (Local Interpretable Model-Agnostic Explanations): This modifies the input of data samples to see how the predictions change. LIME allows the model to be interpreted locally. After altering the feature values, the output is observed.
(2) SHAP (Shapley Additive explanations): It is yet another method for explaining the output of an ML model. It is comparable to LIME in that interpretations are used to show the influence of a certain feature value vs the forecast. SHAP improves a model's transparency.
(3) ELI5 or Explain Like I'm 5: It is a popular Python tool that enables visualisation of ML model predictions and aids in classifier debugging.

ELI5 can be used to evaluate the performance of an ML model in two ways. Global ELI5 demonstrates how parameters behave in relation to the whole model. Local ELI5 examines a single prediction instance.

(4) XAI 5: It is a machine learning toolbox that is specifically built for data analysis and model evaluation. It assists the user in class balance by up-sampling or down-sampling the dataset prior to training and testing. A correlation matrix is used to evaluate the model's behaviour.

(5) Partial Reliance Plot toolkit (PDPbox) is used to compute and visualise the impact of features on target variable prediction. The library is similar to random forest in that it indicates how the characteristic influences prediction. The Python module discovers connections between data or features used as input to the model and the final prediction. It is used to show global and local interpretations of black box models.

(6) GradCAM It allows for the visualisation of heatmaps and gradients for hyper-parameter tuning or the development of confusion matrices, as well as the explanation and visualisation of the prediction.

(7) InterpretML: Microsoft's open-source Python toolbox for training understandable models and explaining black-box technologies. It aids in debugging by allowing you to comprehend the predictions made by a model. It is based on an individual's assessment of global and local aspects.

(8) ALIBI. Alibi is an open-source Python library which consists of a wide range of algorithms to perform interpretation for the prediction at a particular instance.

### 3.3 RQ3: Challenges in the Current Scenario of Trustworthy AI Implementation

The challenges and benefits of using trustworthy AI are listed as follows:

(1) Performance metrics give an insight into the working of a biometric system. Also, they help to compare the different methods used for authentication. However, these metrics, such as accuracy, precision, recall, etc., lack certainty in terms of the decision made by a model. Therefore, trustworthy AI gives a clear reasoning of the decisions taken by the system.

(2) The more confidence in decision-making, the better the model's performance. This increases users' trust and can help give subtle cues in predicting spoofing attacks. Thus, it is essential to ensure that the feedback obtained is accurate.

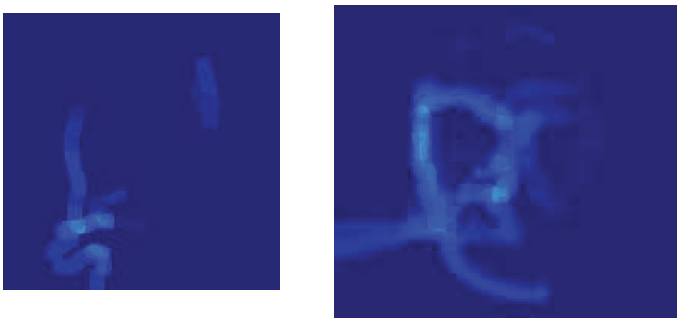

**Figure 6** Heatmaps of faces (a) Real (b) Under attack.

(3) Use of saliency maps or heat maps helps us to acknowledge a model based on the Region of Interest it is looking at. However, when a heatmap is visualized, there might be a difference in opinions of various users-based on the false confidence that a model prediction might lead to. This is because different explainability models mislead users by showing meaningful pixels and leaving them with the task of interpretation.
(4) It is yet to explore if adding a data source can improve the performance of a biometric system, i.e., the inclusion of trustworthiness in a multimodal biometric system, which is another research gap in this field. The inclusion of multiple data sources can support trustworthiness and make them complete.
(5) The aspects that serve as significant background elements for forecasts in terms of bias, security, and privacy become important. Some backdrop features, for example, may appear more frequently on specific predictions. Learning these linked relationships enhances the likelihood that the network will draw conclusions based on them. Demographic data demonstrates that various geographic areas are associated with specific ethnic groupings.

For instance, Figure 6 shows two heatmaps based on the detection of an individual in an original video versus a video under attack.

## 4 Discussions and Future Work

This section discusses the major findings and analysis of the research papers reviewed in trustworthy AI. The section also discusses the impact of the existing research on future avenues based on the findings.

### 4.1 Principal Findings

The main findings can be explained as follows:

1. The most important dimensions of trustworthy AI which have been incorporated in biometric systems are bias & fairness, security & privacy, explainability and performance. 47 research papers have been reviewed in detail for this purpose. It is observed that the maximum amount of research is being done in ensuring fairness in such systems by mitigating bias. This is followed by implementation techniques for security and privacy, followed by focus on explainability of the decisions taken by the system as well as performance of the framework.
2. It is observed that the maximum amount of research has been done in the domain of explainability and bias mitigation in the year 2023. However, the context behind the addition of these trustworthy factors is the improvement of system performance or system security and privacy.
3. Based on our research, there is an overlap in the dimensions addressed in the published work. These are depicted in Figure 7. The dimension highlighted in the paper is mapped with the other dimensions covered in the work.
4. The main advantages of using trustworthiness AI in biometrics are data security and fairness in making decisions. The use of such frameworks ensures security and reduces transaction costs.

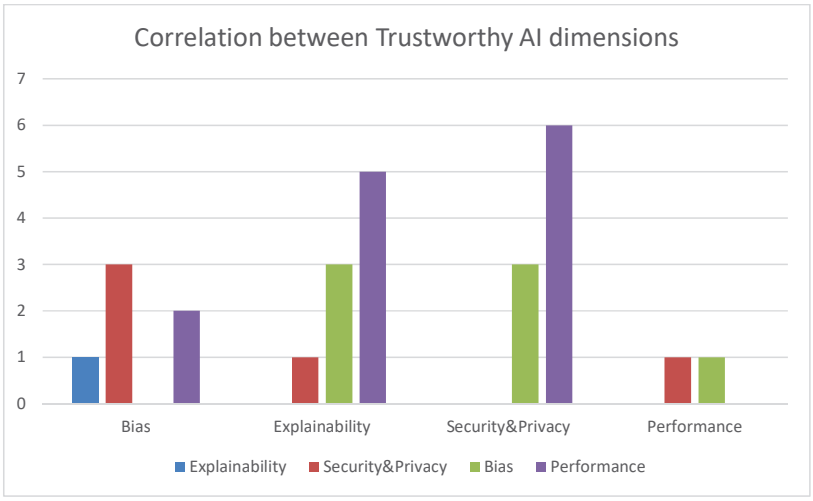


**Figure 7** Correlation of trustworthy AI factors in published research.

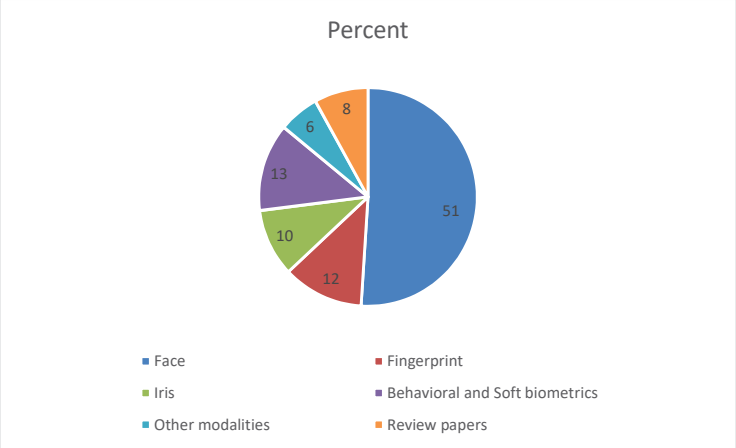


**Figure 8** Break-up of work done in analyzed research papers.

**Table 2** Demography of researchers

| Country | % of Researchers |
|---|---|
| USA | 20% |
| China | 20% |
| Europe | 35% |
| India | 15% |
| UAE & Neighboring Countries | 10% |

5. From the reviewed papers, it can be analyzed that 50% of the current research work is in the domain of facial recognition followed by fingerprints, iris and other modalities. A lot of research is being done in the domain of behavioral biometrics such as motion, keystrokes etc. as well as estimation of soft biometrics such as age, gender. This is shown in Figure 8.
6. The affiliations of researchers are reviewed in Table 2 to understand the amount of research pursued across various countries.

Table 2 depicts the findings with respect to affiliations (approximation of researchers' demographics) as follows:

## 4.2 Impact on Economic Scenario and Policy Framework

When working with biometric systems, the deployment of trustworthy AI ensures human freedom and autonomy [35]. As a result, trustworthiness is

a complicated necessity for people and communities to design, deploy, and operate AI systems, as well as a prerequisite for reaping the potentially enormous societal and economic benefits that AI can deliver. Furthermore, trustworthiness applies not only to the system as a whole, but also to the many players and processes that are involved in it. This necessitates a comprehensive and systemic examination of the pillars and needs that lead to the development of trust in the user of an AI-based system.

## 5 Conclusion and Future Work

The study delves into the fast developing confluence of biometric technology and deep learning, addressing a wide range of trustworthiness-related variables. This detailed review identifies critical areas of research and envisions future effort to improve this potential field based on a thorough examination of existing information. The recommended future courses address a wide variety of concerns and opportunities. The study emphasises the considerable research and frameworks that implement these dimensions in one way or another.

Our findings suggest that a lot of work is being done in the domain of bias and explainability. However, the main aim behind these dimensions is the scaling up of security, privacy and performance of biometric systems. A lot of research involves the facial modality followed by study of fingerprints, iris and behavioral biometrics. The dimensions of trustworthiness are correlated and each study incorporates atleast one of these factors. The review discusses the various frameworks that can be used to add trustworthiness to biometric systems, followed by the benefits and challenges that envelope this domain. In the future, researchers can correlate the dimensions of trustworthiness on several other modalities in which research is ongoing.

## Conflict of Interest

There is no conflict of interest.

## References

[1] Abdullahi, S. B., Khunpanuk, C., Bature, Z. A., Chiroma, H., Pakkaranang, N., Abubakar, A. B., and Ibrahim, A. H. (2022). Biometric information recognition using artificial intelligence algorithms:

A performance comparison. *IEEE Access*, 10, 49167–49183. https://doi.org/10.1109/ACCESS.2022.3179315.

[2] Alshazly, H., Linse, C., Barth, E., Idris, S. A., and Martinetz, T. (2021). Towards explainable ear recognition systems using deep residual networks. *IEEE Access*, 9, 122254–122273. https://doi.org/10.1109/ACCESS.2021.3109899.

[3] Aquino, G., Costa, M. G., and Costa Filho, C. F. (2022). Explaining one-dimensional convolutional models in human activity recognition and biometric identification tasks. *Sensors*, 22(15), 5644. https://doi.org/10.3390/s22155644.

[4] Bowyer, K. W., King, M. C., Scheirer, W. J., and Vangara, K. (2020). The "criminality from face" illusion. *IEEE Transactions on Technology and Society*, 1(4), 175–183. https://doi.org/10.1109/TTS.2020.3013039.

[5] Butt, M. A., Qayyum, A., Ali, H., Al-Fuqaha, A., and Qadir, J. (2023). Towards secure private and trustworthy human-centric embedded machine learning: An emotion-aware facial recognition case study. *Computers & Security*, 125, 103058. https://doi.org/10.1016/j.cose.2022.103058.

[6] Cascone, L., Pero, C., and Proença, H. (2023). Visual and textual explainability for a biometric verification system based on piecewise facial attribute analysis. *Image and Vision Computing*, 132, 104645. https://doi.org/10.1016/j.imavis.2023.104645.

[7] Cavazos, J. G., Phillips, P. J., Castillo, C. D., and O'Toole, A. J. (2020). Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(1), 101–111. https://doi.org/10.1109/TBIOM.2020.2965940.

[8] Chen, Y. Y., Jhong, S. Y., Hsia, C. H., and Hua, K. L. (2021). Explainable AI: A multispectral palm-vein identification system with new augmentation features. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM),* 17(3s), 1–21. https://doi.org/10.1145/3472220.

[9] Curtis, S., Belli, D., Alanoca, S., Bora, A., Miailhe, N., and Lannquist, Y. (2021). Bridging the trust gaps in biometrics. *Biometric Technology Today*, 2021(3), 5–8. https://doi.org/10.1016/S0969-4765(21)00041-4.

[10] Drozdowski, P., Rathgeb, C., Dantcheva, A., Damer, N., and Busch, C. (2020). Demographic bias in biometrics: A survey on an emerging challenge. *IEEE Transactions on Technology and Society*, 1(2), 89–103. https://doi.org/10.1109/TTS.2020.2992349.

[11] Giudici, P., Centurelli, M., and Turchetta, S. (2023). Artificial intelligence risk measurement. *Expert Systems with Applications*, 121220. https://doi.org/10.1016/j.eswa.2022.121220.

[12] Gumaei, A., Sammouda, R., Al-Salman, A. M. S., and Alsanad, A. (2019). Anti-spoofing cloud-based multi-spectral biometric identification system for enterprise security and privacy-preservation. *Journal of Parallel and Distributed Computing*, 124, 27–40. https://doi.org/10.1016/j.jpdc.2018.10.006.

[13] Han, H., and Liu, X. (2022). The challenges of explainable AI in biomedical data science. *BMC Bioinformatics*, 22(12), 1–3. https://doi.org/10.1186/s12859-021-04492-0.

[14] Irshad, A., Usman, M., Chaudhry, S. A., Bashir, A. K., Jolfaei, A., and Srivastava, G. (2020). Fuzzy-in-the-loop-driven low-cost and secure biometric user access to server. *IEEE Transactions on Reliability*, 70(3), 1014–1025. https://doi.org/10.1109/TR.2020.2978458.

[15] Jain, A. K., Deb, D., and Engelsma, J. J. (2021). Biometrics: Trust, but verify. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3), 303–323. https://doi.org/10.1109/TBIOM.2021.3063306.

[16] Jeon, W. S., and Rhee, S. Y. (2017). Fingerprint pattern classification using convolution neural network. *International Journal of Fuzzy Logic and Intelligent Systems*, 17, 170–176. https://doi.org/10.5391/IJFIS.2017.17.3.170.

[17] Kaur, D., Uslu, S., Rittichier, K. J., and Durresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Computing Surveys (CSUR)*, 55(2), 1–38. https://doi.org/10.1145/3494602.

[18] Khairnar, S., Gite, S., Kotecha, K., and Thepade, S. D. (2023). Face liveness detection using artificial intelligence techniques: A systematic literature review and future directions. *Big Data and Cognitive Computing*, 7(1), 37. https://doi.org/10.3390/bdcc7010037.

[19] Khan, R. A., Meyer, A., Konik, H., and Bouakaz, S. (2019). Saliency-based framework for facial expression recognition. *Frontiers of Computer Science*, 13, 183–198. https://doi.org/10.1007/s11704-018-7168-1.

[20] Korgialas, C., Pantraki, E., Bolari, A., Sotiroudi, M., and Kotropoulos, C. (2023). Face aging by explainable conditional adversarial autoencoders. *Journal of Imaging*, 9(5), 96. https://doi.org/10.3390/jimaging9050096.

[21] Lai, K., Oliveira, H. C., Hou, M., Yanushkevich, S. N., and Shmerko, V. P. (2020). Risk, trust, and bias: Causal regulators of biometric-enabled

decision support. *IEEE Access*, 8, 148779–148792. https://doi.org/10.1109/ACCESS.2020.3015710.

[22] Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., and Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys (CSUR),* 55(9), 1–46. https://doi.org/10.1145/3527164.

[23] Lin, Y. S., Liu, Z. Y., Chen, Y. A., Wang, Y. S., Chang, Y. L., and Hsu, W. H. (2021). xCos: An explainable cosine metric for face verification task. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(3s), 1–16. https://doi.org/10.1145/3465027.

[24] Merhav, N. (2018). Ensemble performance of biometric authentication systems based on secret key generation. *IEEE Transactions on Information Theory*, 65(4), 2477–2491. https://doi.org/10.1109/TIT.2018.2877179.

[25] Moraes, T. G., Almeida, E. C., and de Pereira, J. R. L. (2021). Smile, you are being identified! Risks and measures for the use of facial recognition in (semi-) public spaces. *AI and Ethics*, 1(2), 159–172. https://doi.org/10.1007/s43681-020-00018-9.

[26] Nagpal, S., Singh, M., Narain, S., and Bhatnagar, S. (2023). Privacy-enabled biometric authentication using deep learning. *Computers & Security,* 125, 103014. https://doi.org/10.1016/j.cose.2022.103014.

[27] Nanni, L., Brahnam, S., and Lumini, A. (2020). Deep learning for ear biometrics: A survey. *Neurocomputing*, 383, 107–120. https://doi.org/10.1016/j.neucom.2019.11.002.

[28] Nwoye, C. I., and Thompson, K. A. (2020). Facial recognition technology in modern society: Challenges and legal implications. *International Journal of Law and Information Technology*, 28(2), 187–213. https://doi.org/10.1093/ijlit/eaaa010.

[29] Patel, V. M., Smith, L. N., Guerra, L. M., Nasrabadi, N. M., and Chellappa, R. (2019). Automatic target recognition in forward-looking infrared imagery: A survey. *IEEE Access*, 7, 104379–104388. https://doi.org/10.1109/ACCESS.2019.2931598.

[30] Siddiqui, S., Naseer, S., Rana, A., and Afzal, M. (2021). A survey of deep learning techniques for biometric identification. *Journal of King Saud University-Computer and Information Sciences*, 33(10), 1284–1292. https://doi.org/10.1016/j.jksuci.2019.09.005.

[31] Singh, S., Singhal, A., and Jain, A. (2021). Face recognition with mask using deep learning techniques. *Journal of Ambient Intelligence and*

*Humanized Computing*, 12, 9435–9446. https://doi.org/10.1007/s12652-020-02808-8.

[32] Sousa, A., Salami, H., and Soltanalian, M. (2021). Adversarial robustness of biometric recognition: A comprehensive survey. *IEEE Transactions on Artificial Intelligence*, 3(4), 611–628. https://doi.org/10.1109/TAI.2021.3050510.

[33] Uluturk, T. E., Uygun, E., and Yildiz, S. (2023). Face anti-spoofing by using a deep multi-scale neural network architecture. *Neural Computing and Applications*, 35, 5145–5161. https://doi.org/10.1007/s00521-022-08073-4.

[34] Vanarase, T., Koike, K., and Hsu, Y. (2023). Fairness in facial recognition: Do race and gender affect the accuracy of commercial automatic face recognition algorithms? *IEEE Access*, 11, 5719–5736. https://doi.org/10.1109/ACCESS.2022.3234699.

[35] Wang, M., Hu, W., and Zhao, Q. (2022). Privacy-preserving multimodal biometric recognition in cloud environments based on federated learning. *IEEE Transactions on Information Forensics and Security*, 17, 426–438. https://doi.org/10.1109/TIFS.2021.3111711.

[36] Weerasinghe, T. U., Abhayasinghe, N., and Dias, J. (2021). A comparative study on feature extraction approaches for biometric iris recognition using deep learning. *Journal of Imaging*, 7(9), 156. https://doi.org/10.3390/jimaging7090156.

[37] Wu, Y., Wang, R., He, R., and Tan, T. (2022). Robustness and explainability of age and gender recognition in the wild using deep convolutional networks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3), 336–347. https://doi.org/10.1109/TBIOM.2022.3162746.

[38] Yadav, K., and Singh, R. (2023). Biometric recognition using Explainable AI: Opportunities, challenges, and future directions. *Journal of Information Security and Applications*, 74, 103475. https://doi.org/10.1016/j.jisa.2022.103475.

[39] Yin, F., Liu, X., Zhang, X., and Ma, L. (2021). A lightweight convolutional neural network for 2D gait recognition in biometric security systems. *IEEE Access*, 9, 133481–133495. https://doi.org/10.1109/ACCESS.2021.3114609.

[40] Zeng, W., and Liu, H. (2021). Privacy-preserving biometric authentication and verification with secure multiparty computation. *IEEE Transactions on Information Forensics and Security*, 16, 3856–3871. https://doi.org/10.1109/TIFS.2021.3076645.

[41] Zhang, J., Duan, J., and Xu, W. (2022). Biometric recognition using wearable devices: A comprehensive survey. *IEEE Internet of Things Journal*, 9(3), 1561–1579. https://doi.org/10.1109/JIOT.2021.3074683.

[42] Zhao, Z., Zhang, Z., and Zhou, Z. (2021). Towards adversarially robust biometric authentication systems: A survey. *IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(4)*, 351–365. https://doi.org/10.1109/TBIOM.2020.2995995.

[43] Zheng, Y., Huang, Z., and He, Y. (2023). AI-enabled continuous biometric authentication: A review of advances, challenges, and opportunities. *Pattern Recognition Letters*, 164, 129–140. https://doi.org/10.1016/j.patrec.2023.01.012.

[44] Zhou, F., Wu, X., and Zhao, S. (2022). Real-time iris biometric recognition using deep learning-based video sequence processing. *Neural Networks*, 151, 183–195. https://doi.org/10.1016/j.neunet.2022.04.019.

[45] Zou, X., Wang, Y., and Jin, R. (2023). Multimodal biometric authentication via 3D convolutional neural networks and facial gesture recognition. *Pattern Recognition*, 133, 108837. https://doi.org/10.1016/j.patcog.2022.108837.

[46] Zhou, K., and Ren, J. (2018). PassBio: Privacy-preserving user-centric biometric authentication. *IEEE Transactions on Information Forensics and Security*, 13(12), 3050–3063. https://doi.org/10.1109/TIFS.2018.2856975.

[47] Zhu, L., Zhang, C., Xu, C., Liu, X., and Huang, C. (2018). An efficient and privacy-preserving biometric identification scheme in cloud computing. *IEEE Access*, 6, 19025–19033. https://doi.org/10.1109/ACCESS.2018.2820382.

[48] Zhang, H., Li, S., Shi, Y., and Yang, J. (2019). Graph fusion for finger multimodal biometrics. *IEEE Access*, 7, 28607–28615. https://doi.org/10.1109/ACCESS.2019.2899328.

## Biographies

**Shefali Arora** is Assistant Professor in the Department of Computer Science and Engineering at National Institute of Technology Jalandhar. She completed her PhD from Netaji Subhas Institute of Technology Delhi, Masters from Thapar University Patiala and BTech from Indira Gandhi Institute Of Technology Delhi. Here research domain is Deep Learning, Computer Vision and Security. She has published papers in many renowned international conferences and prestigious journals.

**Kanu Goel** is currently working as Assistant Professor in the Department of Computer Science & Engineering, Punjab Engineering College, Chandigarh. She holds a B.E. (Computer Engineering), M.E. (Computer Science and Engineering) and Ph.D., all from Thapar Institute of Engineering & Technology, Patiala. She received a Gold Medal for securing top position in M.E. in Computer Science and Engineering from Thapar University, Patiala in 2016. Her areas of research interests include Machine Learning, Artificial Intelligence, Concept Drift analysis. She has SCIE/Scopus indexed publications in various international journals and conferences.

**Shikha Gupta** has worked as an Assistant Professor at Maharaja Agrasen Institute of Technology Delhi. Her area of expertise is Machine Learning, Data Analytics. She completed her PhD from Netaji Subhas Institute of Technology Delhi.

**Ruchi Mittal** is a seasoned AI/ML professional with over 15 years of experience, specializing in Generative AI, Large Language Models (LLMs), and Natural Language Processing (NLP). With a Ph.D. in Computer Science from the University of Delhi, she has a proven track record of designing, fine-tuning, and deploying advanced AI models, including GPT, BERT, and LLaMA, for applications such as financial forecasting, sentiment analysis, and speech synthesis. Ruchi has published 40+ research papers in SCI, ABDC, and Scopus-indexed journals, focusing on AI/ML, NLP, and computer vision. Currently serving as a Senior Data Scientist-Lead at Iconic Data, Japan, she leads Generative AI projects, optimizes LLMs, and builds scalable AI pipelines. Ruchi is passionate about driving innovation through AI and mentoring teams to achieve technical excellence.

**Avinash Kumar Shrivastava** did B.Sc (H) in Mathematics and received his Master's, M.Phil and PhD degrees in Operational Research from the Department of Operational Research, University of Delhi. His current teaching interest includes courses on decision sciences viz. Business Mathematics & Statistics, Operations Research, Quantitative Techniques, Multi-Criteria Decision Making (MCDM) and Data Analytics. He has presented papers at conferences of International repute and won accolades for best paper presentations. He has been publishing extensively and serving as a reviewer for

various journals of International repute. He has edited seven books published by Bloomsbury Publications & Taylor & Francis. He is the series editor of a book titled "Advances in emerging markets and Business operations". He is the managing editor of the International Journal of System Assurance Engineering and Management (IJSAEM), Springer and associate editor of IMI Konnect. He has organized International conferences & seminars in different capacities. He is also a life member of the Society for Reliability, Engineering, Quality and Operations Management (SREQOM).