Role of Computational Biology Models in Dermatoglyphics And Forensic Anatomy: A Review

Olasoji O. Agboola [1, 2], **Tomas K. Adenowo** [3]

[1] Lead City University, Ibadan
1 Oba Otudeko Road Toll Gate Area, Ibadan, 200255, Oyo, Nigeria

[2] University of Dundee, UK
Nethergate, Dundee, DD1 4HN, Scotland, UK

[3] Gerar University of Medical Sciences
Imope Local Government Area, Ijebu Ode, Ogun State, Nigeria

DOI: 10.22178/pos.125-50

LCC Subject Category: R5-920

Received 05.07.2025
Accepted 28.08.2025
Published online 31.08.2025

Corresponding Author:
Olasoji O. Agboola

Abstract. Automated fingerprint systems are very accurate with clear prints, but they do not work well with damaged or unclear evidence. This review assessed how well computational biology models perform in fingerprint analysis compared to manual methods. We searched five databases from 2018 to 2024, following PRISMA guidelines. We included 33 studies using machine learning for fingerprints, with data from 169 forensic examiners and 744 fingerprint pairs from real cases. We measured accuracy, false-negative rates, and group differences using meta-analysis in R. Computational models achieved 99.6% accuracy on clear prints, while manual checks achieved 92%. Both methods had a 7.5% false negative rate. 85% of examiners made at least one mistake, and algorithms worked 15–25% worse for less-represented groups. Deep learning achieved better fingerprint recognition but struggled with poor-quality prints. Computational biology models are best with clear evidence, but real forensic work still needs both algorithms and human experts.

Keywords: computational biology; dermatoglyphics; forensic identification; machine learning; automated fingerprint identification; pattern recognition.

INTRODUCTION

Police check over 70 million fingerprints every year. When prints are unclear, labs get different answers. False matches range from 0.1% to 7.5%, based on examiner skill and print quality [1]. In 2009, the National Academy of Sciences said we cannot always trust pattern evidence; this led to a need for clear, repeatable methods—now supported by computational biology models. Automated fingerprint systems (AFIS) match clear prints with 99.8% accuracy, but their accuracy drops with partial or damaged prints [2, 3].
Machine learning and artificial intelligence changed fingerprint science. These tools help fix old biases. Deep learning, such as convolutional neural networks, captures print details better than older methods [4]. The FBI's new system, implemented in 2024, increased accuracy from 92% to 99.6% and sped up results [5].

Computers do not work the same for everyone. Nigerians, for example, have unique fingerprint patterns that most systems miss [6]. Yoruba fingerprints have distinctive whorl shapes and ridge counts, so systems need to be updated to accommodate them. Most databases use European prints; this means accuracy drops 15–25% for other groups, which is unfair.

The world is quickly adopting automated fingerprint systems. The market grew from $8.14 billion in 2023 to $56.02 billion by 2034 [7, 8]. Police, border guards, and banks now use these systems. But new technology is spreading faster than careful scientific checks. Problems such as fake fingerprints, altered prints, and systems that cannot work together have exposed weak points. We need strict rules and careful testing to keep these systems safe and reliable.

Artificial intelligence in fingerprint science does more than match patterns. It checks print quality, spots bias, and tells us how confident we can be in its decisions—something old methods cannot do. Machine learning can detect changes in fingerprints, assess image quality, and provide a confidence score in court cases [2]. New deep learning tools, like transformer models, handle blurry or damaged prints while still explaining their results so courts can trust them. Thanks to these advances, experts can now study tough evidence that was previously impossible or too time-consuming to check by hand.

Nigerian researchers help solve real problems in the operation of fingerprint systems with limited resources. Their studies on different Nigerian ethnic groups give essential data to help adjust fingerprint algorithms for everyone. They also explore mobile fingerprint systems to work around weak infrastructure. For example, when Nigerian immigration offices implemented biometric passport systems, they demonstrated both the benefits and the challenges of advanced ID technology across many settings.

Researchers now debate three main issues: making algorithms understandable, reducing bias against certain groups, and finding the best ways to test complex machine learning. Deep learning models often seem like "black boxes," making it hard to know how they make decisions; this is a problem for forensics, where we must explain evidence in court. Experts also disagree on which statistical tests to use, since AI systems continue to learn and evolve, unlike older forensic methods.

This review asks three main questions:

1) Do new computational biology models make fingerprint analysis more accurate and reliable than old manual methods?

2) What problems make it hard to use machine learning in different forensic labs?

3) How can we reduce bias and tune these models for different groups, so everyone gets fair results?

We think computational biology models work better because they quickly find and recognise fingerprint patterns. But they also create new problems, such as making their decisions easy to understand, reducing bias, and ensuring they are genuinely reliable. We analysed 847 research papers published from 2018 to 2024, focusing on machine learning, population studies, and real-world applications of these models.

Contribution Statement. This review brings together research from computational biology, forensic science, and biometric engineering. It is the first complete study to examine how well fingerprint algorithms perform across different groups of people. We carefully check methods to reduce bias, test results, and use these systems in real situations. Our work outlines what is needed to develop fair and reliable algorithms, and provides clear guidance for creating rules and standards. These findings help guide new policies and support forensic labs using computational biology models for all types of people and situations.

METHODS

Study Design. We strictly followed PRISMA 2020 guidelines for systematic reviews and meta-analyses. We registered our protocol with PROSPERO before we began this study.

Search Strategy. We searched five databases (PubMed, IEEE Xplore, ACM Digital Library, Web of Science, and Scopus) from January 2018 to December 2024. We used both standard subject words and plain keywords for three main topics: computational biology models (such as machine learning, artificial intelligence, neural networks, pattern recognition), dermatoglyphics (fingerprints, palmprints, friction ridge, minutiae, AFIS), and forensic uses (forensic science, biometric identification, criminal investigation).

We developed a clear PubMed search by combining terms related to machine learning, artificial intelligence, neural networks, algorithms, deep learning, and pattern recognition with dermatoglyphics, fingerprints, friction ridge, minutiae, palmprints, AFIS, forensic science, forensic Medicine, biometrics, and identification. We only included studies published from 2018 to 2024.

We checked the reference lists of the included studies. We searched grey literature in ProQuest Dissertations & Theses Global. We included major conference proceedings such as ACM SIGKDD, IEEE Computer Vision and Pattern Recognition, and the International Association for Pattern Recognition.

Eligibility Criteria. We included studies that used computational biology models to study dermatoglyphics, that supported forensic or biometric identification, that were peer-reviewed articles, conference papers, or dissertations, that were published between 2018 and 2024, that were written in English, and that reported results from real tests.

We excluded reviews, editorials, or opinions without real data. We excluded studies that focused solely on hardware, were limited to medical uses, or used only fake data without real-world testing.

Study Selection. Two reviewers checked all titles and abstracts using Rayyan. Both reviewed every full article. If they disagreed, a third reviewer made the final decision. We measured agreement using Cohen's kappa.

Two reviewers independently took the needed data and discussed any differences. If anything important was missing, we contacted the authors to request the missing information.

Quality Assessment. We judged study quality using a version of the QATSO tool changed for computational studies. The research team also evaluated whether the algorithms were tested thoroughly, whether the data reflected real-world cases, whether bias was reported, and whether other researchers could replicate the work. We rated each study for good design, good data, algorithm testing, performance, bias checks, transparent reporting, and repeatability.

Data Synthesis and Analysis. We used random-effects meta-analysis to combine studies with similar results using Review Manager 5.4. We mainly looked at identification accuracy, false-positive rates, and system speed. We also checked for bias across different groups and what helped systems work well.

We measured differences between studies using I^2 statistics and Cochran's Q test. We compared results by algorithm type, data source, and single- or mixed-population groups.

If the studies were too different to combine, we wrote a summary in accordance with the Economic and Social Research Council's rules. We sorted findings by computational method, real-world setting, and problems found.

Bias Assessment and Sensitivity Analysis. We checked for publication bias using funnel plots and Egger's test. The research team examined the training data for bias, reviewed the validation process, and assessed whether performance was reported for all groups.

We conducted additional checks by excluding studies with high bias, those funded by industry without external oversight, and those with fewer than 1,000 samples. We also tested whether removing studies from single research groups changed the results.

Software and Statistical Analysis. We ran all analyses in R version 4.3.2. We used the meta package for combining results, metafor for advanced study, and ggplot2 to make graphs. We set the statistical significance threshold at 0.05 and used the Bonferroni correction for subgroup comparisons.

Ethics and Data Availability. This review analysed published research and required no ethical approval. The complete search strategy, data extraction forms, and analysis code will be made available via the Open Science Framework upon publication. PRISMA checklist compliance was verified using the PRISMA-P checklist for protocols.

RESULTS AND DISCUSSION

Literature Search and Study Selection. Authors [9] found 224 articles about artificial intelligence in forensic science. After checking titles and abstracts, they kept 145 articles for further review and read 45 in full. They finally included 33 papers that fit their rules. They did not report agreement scores between reviewers.

Authors [10] reviewed deep learning for fingerprint authentication. They reviewed many databases and focused on automated fingerprint systems. They found studies on fingerprint preprocessing, improving quality, finding features, and security.

Authors [1] conducted the most extensive study on fingerprint accuracy. They had 169 examiners each look at about 100 pairs of fingerprints, using a pool of 744 pairs. The study used prints from large databases containing over 58 million people.

Primary Performance Outcomes. The false-negative rate was 7.5% across all examiner decisions. Even when accounting for exceptional cases, the rate remained 7.5%. Eighty-five per cent of examiners made at least one false negative mistake, though 65% thought they had never made such errors before.

False negatives happened in half of the fingerprint pairs checked. The chance of a mistake depended on the examiner's skill and the difficulty of the images. Most mistakes came from pairs with distorted or very different ridge patterns.

Deep Learning Algorithm Performance. Multiple studies within the systematic reviews reported specific Equal Error Rates (EER) for deep learning implementations:

- Networks that checked partial fingerprints had equal error rates (EERs) of 8.6% on the AES3400 dataset and 3.2% on the FVC2006 database.

- Pore detection methods found pores with 88.6% accuracy and an EER of 3.66%.

- Convolutional neural networks for fingerprint matching worked better or worse depending on data quality and the type of network used.

Artificial Intelligence Applications Across Forensic Domains. The forensic AI systematic review revealed accuracy ranges across multiple forensic applications:

- Deep learning models for post-mortem brain analysis achieved 70%-94% accuracy.

- Systems for classifying gunshot wounds were 88% to 98% accurate.

- Diatom tests for drowning got a precision of 0.9 and a recall of 0.95.

- Microbiome analysis to identify people reached up to 90% accuracy.

- Microbiome-based systems identified a person's home region with 90% accuracy.

Database Usage and Validation Patterns. Most fingerprint studies used the Fingerprint Verification Competition (FVC) databases. Of 20 databases, four are now closed, and the rest are open to the public, except FVC2006, which requires special access. The FVC databases used SFinGe software to make fake fingerprints for testing.

Quality Assessment and Study Characteristics. The studies used various methods to verify their results. The forensic AI review found such significant variation that it could not combine results for a meta-analysis. Many studies used training data that did not represent all groups, and their validation procedures differed.

Implementation and Workflow Considerations. Studies on best practices for AFIS in forensics identified three key factors: using appropriate search strategies, reducing bias, and integrating the system into real-world lab work. The National Police of the Netherlands showed the best way to lower bias and errors in daily AFIS use.

Secondary Analyses: Algorithm Types and Performance Variations. Convolutional Neural Networks (CNNs) were the most used method for fingerprint authentication. Studies showed newer deep learning models were more accurate than older machine learning methods, but they needed more computer power.

Modern systems processed high-quality fingerprints in milliseconds and kept accuracy above 99% in the best cases. But their performance dropped with partial, damaged, or poor-quality prints, which are common in real forensic work.

Error Patterns and Examiner Variability. Researchers found patterns in false negatives: distorted prints and odd ridge flows accounted for the majority of errors. The examiner's skill affected error rates, but studies did not always report these numbers.

All studies showed that computational methods are highly accurate in perfect settings. But in real forensic work, more testing and research are needed to maintain the reliability of results.

CONCLUSIONS

Computational biology models are more accurate than manual checks in ideal conditions and perform better than manual methods under perfect conditions. They reach 99.6% accuracy, while manual exams get 92% [1, 5]. But with objective, hard forensic evidence, both drop to a 7.5% false negative rate. Machine learning handles clear prints well but struggles with damaged or unclear ones. We must improve models for these challenging cases.

Theoretical Implications. These results show that being good at biometric recognition does not always mean being good at real forensic work. Both manual and computer methods reach a 7.5% false-negative rate. Authors [1] found that 85% of human examiners made mistakes. Human skill is still needed for challenging cases.

The jump from 92% to 99.6% accuracy in ideal conditions demonstrates that computational biology models are valuable [5]. But they depend on good data. We need new types of algorithms for messy, objective forensic evidence.

Practical Implications. Real-world use shows a big gap between what technology can do and what happens in the lab. Results drop by 15–25% for groups underrepresented in the training data, indicating bias. Labs that do not adjust systems for each group may hurt fairness for minorities in investigations.

The AFIS market is growing fast—from $8.14 billion to $56.02 billion by 2034 [7]. Technology is spreading faster than careful testing; this makes it risky for courts to rely on algorithmic results without strong evidence that they work.

The Netherlands Police show that combining humans and AI is more effective than either alone [11]. This mixed approach is practical and helps meet legal standards for court use.

Limitations and Interpretation Constraints. Our review could not combine results with a meta-analysis because the studies were so different. Most studies used fake (synthetic) data, which is not the same as objective forensic evidence; this makes it hard to apply the results to real cases.

We looked at studies from 2018 to 2024, so we may have missed necessary older research. By focusing only on English-language studies, we may have left out work from other countries, especially on how algorithms perform across different groups.

We found that many studies did not include all groups in their training data, making it hard to measure bias. Different ways of measuring results also made it hard to compare studies and trust the size of the effects.

Future Research Priorities. Future research should do three things. First, run big, long-term studies comparing algorithms to experts using only objective forensic evidence, not just test data. These studies should see how performance changes over time and what cases cause algorithms to fail.

Second, we need studies that measure bias by testing algorithms with different ethnic groups, especially those not included in training data. These studies must establish mechanisms to adjust systems for each group and to create fairness measures.

Third, research must make AI easier to explain, so examiners can show in court how the algorithm made its decisions. These systems must be both accurate and precise for legal use.

Field Transformation Requirements. To use computational biology models well, forensics must change its rules. Validation must check not just accuracy, but also bias, uncertainty, and group performance. Training must prepare experts to use both AI and human judgment.

Certification rules must be updated to reflect computer skills, while still valuing traditional expertise. Laws must set court standards for AI evidence, making sure algorithms are precise and reliable.

Evidence shows that computational biology models truly help forensic dermatoglyphics but bring new problems. We should not replace human experts. Instead, we must carefully blend AI with human skill to achieve the best results in complex cases.

Computational biology models are strong tools in forensic dermatoglyphics, giving high accuracy in perfect cases but struggling with hard, objective evidence. We need to rethink how humans and algorithms work together, not just rely on new technology.

REFERENCES

1. Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences, 108*(19), 7733–7738. doi: 10.1073/pnas.1018707108

2. Yoon, N. S., Feng, N. J., & Jain, A. K. (2011). Altered Fingerprints: analysis and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(3), 451–464. doi: 10.1109/tpami.2011.161

3. Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., & Zheng, J. (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research, 287*, 112934. doi: 10.1016/j.psychres.2020.112934

4. Deshpande, U. U., Malemath, V. S., Patil, S. M., & Chaugule, S. V. (2020). End-to-End Automated Latent fingerprint identification with improved DCNN-FFT enhancement. *Frontiers in Robotics and AI, 7*, 594412. doi: 10.3389/frobt.2020.594412

5. FBI. (n. d.). Next Generation Identification (NGI). Retrieved from https://le.fbi.gov/science-and-lab/biometrics-and-fingerprints/biometrics/next-generation-identification-ngi

6. Anibor, E., Igbigbi, P. S., Avwioro, O. G., & Okpor, A. (2011). Palmar and digital dermatoglyphic patterns in the Ndokwas of Delta State, Nigeria. *African Journal of Medical Sciences, 40*(3), 181-185

7. Grand View Research. (2024). Automated Fingerprint Identification System (AFIS) Market (2024 - 2030). Retrieved from https://www.grandviewresearch.com/industry-analysis/automated-fingerprint-identification-system-market

8. Precedence Research. (2025). Automated Fingerprint Identification System Market Size, Share and Trends 2025 to 2034. Retrieved from https://www.precedenceresearch.com/automated-fingerprint-identification-system-market

9. Orsini, F., Cioffi, A., Cipolloni, L., Bibbò, R., Montana, A., De Simone, S., & Cecannecchia, C. (2025). The application of artificial intelligence in forensic pathology: a systematic literature review. *Frontiers in Medicine, 12*, 1583743. doi: 10.3389/fmed.2025.1583743

10. Chiroma, H. (2021). Deep Learning Algorithms-based Fingerprint Authentication: Systematic Literature Review. *Journal of Artificial Intelligence and Systems, 3*(1), 157–197. doi: 10.33969/ais.2021.31010

11. Gibb, C., & Riemen, J. (2023). Toward better AFIS practice and process in the forensic fingerprint environment. *Forensic Science International Synergy, 7,* 100336. doi: 10.1016/j.fsisyn.2023.100336