MSC Artificial Intelligence: Research Methods and Professional Practice

Unit 7

IMPLEMENTING MACHINE LEARNING TOOLS AND/OR TECHNIQUES IN SUSPECT PROFILING

Yassir Ibararh

University Of Essex

16 June 2025

Introduction:

Crime is a phenomenon that affects all modern societies and poses constant challenges to law enforcement agencies around the globe such as the local police, FBI or the CIA (Gregory, 2005). In the past few years, there has been an increase in the number of crime cases reported due to an escalation in the utilisation of technology, both by the authorities trying to detect and investigate them, and the offenders committing the crimes (Tajuddin et al. 2019).

Until recently, the standard approach to profiling suspects was relying heavily on manual investigation, with human instinct, eyewitness testimony and the physical evidence collected (Alison et al. 2007). The rise in the amount of data has caused the agencies to face challenges when dealing with large volumes needed to be analysed for suspect profiling and identification (Gregory, 2005).

In response to evolving challenges, and the rapid development of AI technologies, advanced suspect profiling techniques such as criminal profiling, facial recognition and behavioural analytics based on machine learning are transforming the way suspects are identified and tracked. As a result, many law enforcement agencies have considered implementing these technologies. (Sachoulidou, 2023).

Although the implementation can help overcome many challenges such as manual data curation and analysis, and time consumption, the researchers are still sceptical about it due to many possible downsides that can be either technical, ethical or legal (Sachoulidou, 2023). The continuous advancement of technology has led many agencies to consider implementing AI-driven profiling tools.

This literature review adopts a structured approach and will focus on examining the types of traditional suspect profiling systems and techniques. It will include an evaluation of peer-reviewed research papers on the implementation of machine learning tools and techniques into the field and will discuss the future effects and possible improvement areas. The data extraction process is focused on recent publications of from the past 6 years with a maximum span of 20 years old. Different types of literature were included such as case studies, journal articles and research papers.

Main Discussion

BinDarwish et al. (2020) define suspect profiling systems as analytical techniques or tools that are primarily used in law enforcement and criminal investigations. They enable the investigators to identify possible suspects based on multiple characteristics and evidence that can be behavioural, demographic, psychological and sometimes even physical. These systems help narrow down investigations by generating profiles for offenders, linking related cases, and prioritising suspects based on historical data and crime patterns.

According to Saeed (2014), there are a couple examples of current profiling systems in suspect identifications, which notably differ based on the intended use case. One of the examples is the VICAP (Violent Criminal Apprehension Program). Research made by Kocsis (2020) emphasises that the tool was developed by the FBI to help law enforcement agencies to identify and link patterns in violent crimes. VICAP collects and analyses detailed data on historical incidents, which are linked to detect any similarities or anomalies.

As a result, the tool has proven to be efficient in improving investigations efficiency and enhanced the ability of detecting possible perpetrators.

Other systems such as AFIS (Automated Fingerprint Identification systems), or IBIS (Integrated Ballistic identification system) are also well known within the law enforcement agencies. Each one is used based on the purpose of analysing a certain specific data (Viorel, 2020). Moreover, a key data source used by such systems is the CODIS (Combined DNA Index System). It is a national DNA database operated mainly by the FBI. It allows for comparison of DNA profiles collected from the crime scenes with those of registered offenders (Viorel, 2024).

Saeed (2014) describes the importance of suspect profiling systems from the concept of focused investigations by helping prioritising suspects based on a scientifically informed criterion. Building on this, Kocsis (2020) confirms their efficiency in saving investigation time by narrowing down leads and using a data-driven approach to take informed decisions. Using such systems enables faster identification and helps improve the public safety by assisting the law enforcement agencies in catching dangerous offenders.

Nevertheless, despite these advantages, traditional suspect profiling methods are not without limitations. According to Shendel (2019) current profiling methods often rely heavily on subjective interpretations, in turn can introduce bias and limit the reliability of such systems across diverse cases. Furthermore, they can fail to produce significant results when applied to offenders that do not fit a typical demographic pattern or a specific behaviour.

Shendel (2019) also contends that the suspect profiling systems are time consuming when requiring manual data-analysis and expert judgement. This limitation becomes particularly problematic in real-time scenarios, where urgency and scale challenge traditional methods. As a result, many agencies are turning to machine learning to augment or replace these methods.

Machine Learning is a branch of artificial intelligence that allows computer systems to learn from data and improve their accuracy and performance on specific tasks without being necessarily programmed or supervised by a human being (Khanzod, et al. 2020). Such systems learn from recurring patterns in the data to make decisions or predictions by following various learning methods.

These methods can differ based on the purpose and the nature of the data being used. The main types include supervised learning, where models are trained on labelled data, and unsupervised learning, which finds patterns in unlabelled data. Additionally, reinforcement learning, where agents learn by interacting with an environment through rewards and penalties, and semi-supervised learning, which combines small amounts of labelled data with large volumes of unlabelled data to improve learning efficiency (Khanzod, et al. 2020).

On one hand, Rich (2016) argued that implementing machine learning in suspect profiling systems can improve their general speed and increase efficiency by automating data analysis and reducing investigation time. The study made by the author highlighted how automating pattern recognition and suspect matching can significantly reduce the workload on investigators and allow them to focus more on strategy rather than manual tasks.

On the other hand, Kuzmanov (2025) shifts focus from efficiency to proactivity. He emphasises that reducing analysis time can help in crime prevention by identifying dangerous criminals in a shorter period. His research showed that early detection using predictive algorithms can assist officers in making timely interventions and preventing crimes before they escalate. Kuzmanov (2025) also argued that in fast-paced environments such as urban areas with high crime rates, having a real-time profiling mechanism can give law enforcement a much-needed edge in narrowing down suspects quickly.

However, the author also warns that such systems should be regularly audited to ensure the results remain fair and unbiased, especially in communities where the historical data might already reflect inequalities or flawed practices. In essence, both Rich (2016) and Kuzmanov (2025) recognise the benefits of machine learning in profiling, but their emphasis diverges—Rich (2016) focuses on reactive efficiency, while Kuzmanov (2025) stresses proactive prevention.

Pieringer et al. (2021) conducted a case study on the application of machine learning in policing in Chile, utilising historical arrest data from the population of Santiago. The set contained records of 777,724 arrests involving 332,602 individuals. The main goal was to evaluate the feasibility of implementing machine learning algorithms in forecasting the likelihood of possible arrests and predicting perpetrators. Using multiple techniques such as group-based clustering which employs K-means algorithm, the results showcased great potential by affirming most of the true positives and false negatives. Nonetheless, there was still a significant number of false positives (wrongfully profiled) with a percentage of 5.8%. If left unaddressed, such inaccuracies remain a major limitation that could severely affect real-world decision-making.

Kumar et al. (2018) outlined some of the current applications of machine learning in suspect profiling. One of the areas mentioned was the facial recognition and identification. Machine learning models are carefully trained on very large datasets such as MegaFace or MS-Celeb-1M that contains thousands of images of previous criminals and registered citizens.

Another key implementation is in predictive policing, where machine learning algorithms analyse historical data of crimes to forecast future events that might be likely to occur (Sajid, 2019). Tools like PredPol or HunchLab use past crime spots, time, and types to predict what is called 'hotspot maps' this can successfully guide the police patrols to areas where danger might occur.

A third growing application is behavioural analysis. Machine learning is used here to detect patterns in the suspect's social or psychological behaviour, often through an analysis of digital footprints. These can be either social media activity, search or messages history (Bhatt et al. 2023). The FBI's Behavioural Analysis Unit (BAU) is well known of using such techniques to enhance traditional suspect profiling. According to Bhatt et al. (2023), models trained on behavioural data have shown up to 72% accuracy in matching suspect traits across similar cases which makes them highly efficient.

According to BinDarwish (2023) machine learning models can cause ethical issues as well. They may inherit bias from inefficient training data, causing unfair profiling of certain groups. This can especially affect minorities or underrepresented populations in the dataset.

Another key issue is privacy. Although facial recognition and image analysis of social media content can help track suspect movements and affiliations, these practices may also raise serious privacy concerns. Without proper consent, they can be perceived as invasive or even unconstitutional (BinDarwish, 2023).

Although machine learning systems may appear highly efficient and accurate, there is often the slight risk of false positives. These are cases where innocent individuals are mistakenly flagged as suspects. Therefore, an overreliance on predictions may lead into wrongful convictions or even a slight suspicion might defer the attention of the investigators from the real suspects and dangerous criminals (Shendel, 2019).

While the applications of machine learning in suspect profiling show promising results, several gaps require a call for further research to maximise the benefit of using such systems in the future. As mentioned by Haley et al. (2025), current systems often suffer from limited generalisation due to incomplete or biased training data. This can compromise both the accuracy and fairness of the results. Further studies are encouraged to explore more inclusive datasets and employ robust training methods to minimise such issues.

One of the solutions to address the current limitations that was proposed by Antonov et al. (2021) is to incorporate explainable AI (XAI). This technique can help make the machine learning decisions more transparent to investigators, which aligns with the current regulations that require interpretability and accountability in high-stakes decision making situations.

Emerging technologies such as deep learning, edge AI, and multimodal profiling (e.g., combining voice, facial expressions, and behaviour) will likely reshape future capabilities (Antonov et al. 2021). The integration of real-time analytics and decentralized data sharing platforms can further enhance responsiveness and cross-agency collaboration. These advancements present significant opportunities to make profiling more accurate, efficient, and ethically sound, if implemented with careful consideration.

Conclusion:

To conclude, machine learning can significantly enhance suspect profiling by enabling faster, more accurate, and data-informed criminal investigations. The existing literature has proven the positive impact that can be achieved when such tools are properly trained and deployed across a wide range of profiling tasks. However, its implementation must be transparent, ethical, and legally regulated to avoid misuse and ensure justice. Without proper safeguards, there is a real risk of reinforcing existing biases or causing harm through false positives.

Generally, the future seems to be bright and full of potential. Proper regulations and considerations must be put in place to ensure a seamless integration and to avoid any negative impacts. Further research is needed to explore how machine learning models can be made more explainable and fairer, especially when applied in high-stakes environments like law enforcement. Collaboration between developers, legal experts, and investigators will be essential going forward. With the right balance between innovation and accountability, suspect profiling systems driven by AI can truly become an asset in modern policing without compromising ethics or public trust.

References:

Alison, L., McLean, C. and Almond, L. (2007) ‘Profiling suspects’, *in Applied Psychology: Research, Training and Practice*. Edited by D. Carson and R. Bull. 2nd edn. London: Routledge, pp. 367-384. Available at: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203118177-24/profiling-suspects-laurence-alison-clare-mclean-louise-almond (Accessed 6 June 2025).

Bhatt, P., Sethi, A., Tasgaonkar, V., Shroff, J., Pendharkar, I., Desai, A. (2023) Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions, Brain Informatics, 10(18), pp. 5-10. Available at: https://link.springer.com/article/10.1186/s40708-023-00196-6 (Accessed 16 June 2025).

BinDarwish, A., Alhammadi, S., Salehi, A. (2023) ‘Crime Detection and Suspect Identification System ‘, *Computer Information Science*, 11(9), pp. 1-11. Available at: https://www.preprints.org/frontend/manuscript/c5c48b6a1b404deb0bf5bea7dd321052/download_pub (Accessed 7 June 2025).

Gregory, N. (2005) ‘Offender profiling: a review of the literature‘, *The British Journal of Forensic Practice*, 7(3), pp. 29-34. Available at: https://doi.org/10.1108/14636646200500019 (Accessed 1 June 2025).

Haley, P., Burrell, D. (2025) Using Artificial Intelligence in Law Enforcement and Policing to Improve Public Health and Safety. Law, Economics and Society. 1(1), pp. 46-47. Available at: https://j.ideasspread.org/index.php/les/article/view/1448 (Accessed 16 June 2025).

Kocsis, R. (2015) ‘The Name of the Rose and Criminal Profiling: The Benefits of ViCAP and ViCLAS ‘, *Journal of Forensic Psychology Practice*, 15(1), pp. 58-79. Available at: https://www.tandfonline.com/doi/abs/10.1080/15228932.2014.986706 (Accessed 7 June 2025).

Khanzod, A., Chhaya, K (2020) ‘ADVANTAGES AND DISADVANTAGES OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING: A LITERATURE REVIEW ‘, *International Journal of Library & Information Science* (IJLIS), 9(1), pp. 30-36. Available at: http://www.iaeme.com/IJLIS/issues.asp?JType=IJLIS&VType=9&IType=1 (Accessed 10 June 2025).

Kuzmanov, I. (2025) ‘The Mind Hunt: Criminal Profiling Before Artificial Intelligence‘, *Criminology and Psychopathology at London Metropolitan University*, 3(1), pp. 1-15. Available at: https://www.researchgate.net/profile/Iliyan-Kuzmanov/publication/388155835_The_Mind_Hunt_Criminal_Profiling_Before_Artificial_Intelligence/links/678c150475d4ab477e4d4389/The-Mind-

Hunt-Criminal-Profiling-Before-Artificial-Intelligence.pdf (Accessed 14 June 2025).

Kumar, A., Malathi, S., Vengatesan, K., Ramakrishan, M (2018) Facial Recognition System for Suspect Identification Using a Surveillance Camera, Pattern Recognition and Image Analysis, 28(1), pp. 410-420. Available at: https://link.springer.com/article/10.1134/S1054661818030136 (Accessed 15 June 2025).

Pieringer, C., Torres, D., Assahi, K. (2021) 'Machine learning for policing: a case study on arrests in Chile ', *An International Journal of Research and Policy*, 31(9), pp. 1036-1050. Available at: https://www.tandfonline.com/doi/full/10.1080/10439463.2020.1779270?needAccess=true#d1e198 (Accessed 12 June 2025).

Rich, M. (2016) 'MACHINE LEARNING, AUTOMATED SUSPICION ALGORITHMS, AND THE FOURTH AMENDMENT ', *University of Pennsylvania Law Review*, 164(4), pp. 871-929. Available at: https://www.jstor.org/stable/24753528 (Accessed 11 June 2025).

Shendel, S. (2019) 'The Challenges of Risk Profiling Used by Law Enforcement: Examining the Cases of COMPAS and SyRI', *in Regulating New Technologies in Uncertain Times*. Hauppauge: T.M.C. Asser Press, The Hague, pp. 225-240. Available at:

https://link.springer.com/chapter/10.1007/978-94-6265-279-8_12 (Accessed 9 June 2025).

Sachoulidou, A. (2023) ‘Going beyond the “common suspects”: to be presumed innocent in the era of algorithms, big data and artificial intelligence’, *Artificial Intelligence and Law*, 10(1), pp. 15-21. Available at: https://link.springer.com/article/10.1007/s10506-023-09347-w (Accessed 10 June 2025).

Saeed, U., Usmani, A., Mikhtar, A. (2015) ‘Application of Machine learning Algorithms in Crime Classification and Classification Rule Mining‘, *Research Journal of Recent Sciences*, 4(3), pp. 106-114. Available at: https://www.isca.me/rjrs/archive/v4/i3/15.ISCA-RJRS-2013-1005.pdf (Accessed 13 June 2025).

Sajid, A., Sheeraz, A., Alvi, A., Rehman, M (2019) ‘The Usual Suspects: Machine Learning Based Predictive Policing for Criminal Identification ‘, 2019 13th *International Conference on Open-Source Systems and Technologies (ICOSST)*, Lahore, Pakistan, 2019, pp. 1-6. Available at: https://ieeexplore.ieee.org/abstract/document/9043925 (Accessed 15 June 2025).

Tajuddin, T., Abd Manaf, A., Awang, F., Dawam, R. (2019) 'Crime Suspect Profiling (CSP) for Forensic Investigation on Smartphone', *2019 4th International Conference and Workshops on Recent Advances and*

Innovations in Engineering (ICRAIE). Kedah, Malaysia, 27-29 November. 127-136. Available at: https://ieeexplore.ieee.org/abstract/document/9037772 (Accessed 5 June 2025).

Viorel, V. (2024) 'THE ROLE OF FORENSIC DATABASES IN INTERNATIONAL COOPERATION ', *Romanian Journal of Forensic Science*, 25(138), pp. 161-165. Available at: https://www.proquest.com/openview/fbb60f8362d7947a6708ac8a608beff0/1?pq-origsite=gscholar&cbl=756356 (Accessed 8 June 2025).