

The Intersection of Genetics and Machine Learning: Utilizing Next-Generation DNA Phenotyping to Reconstruct Criminal Suspect Faces in Forensic Investigations

Author: Mohammad Mostaan Abedi

Institution: Herat Afghan-Turk Boys High School, Herat, Afghanistan

Date: September 2026

Imagine that technology has reached its peak, and with the latest forensic equipment, you can solve any murder case. At that very moment, a murder takes place, and you enter the crime scene with complete confidence that you can easily solve the case. You have already reassured all the victim's family members that, in the shortest possible time, you will identify the killer and send him to prison.

But when you enter the crime scene, you notice something strange; there are no fingerprints, no murder weapon, and no evidence whatsoever indicating who the killer was. But suddenly, you see a drop of blood on the wall, and you become happy because you are going to identify the criminal using this very bloodstain.

But when you examine the drop of blood, you find nothing. Why? Because the DNA is not present in your database, and the killer has never committed any crime before. Disappointed, you are forced to close the case.

But now it is 2026, and you are once again faced with the same crime scene; no murder weapon, no fingerprints, and no evidence of the killer, except for that same bloodstain.

And this time, once again, you become happy when you find the bloodstain, but this time your happiness is not temporary; because now you have a technology that, with just one drop of blood and DNA, can determine the killer's hair color, eye color, skin color, approximate age, and even facial features, allowing you to easily identify the killer.

The name of this astonishing technology that makes thousands of families happy knowing that the killer of their loved ones will receive the punishment they deserve is **DNA Phenotyping**.

As a student living in a country where crime and murder are something normal and ordinary, and where 1900s technology has been used for years to solve murder cases, I can say with confidence that getting away with murder here is much easier than writing a two-page paper. Even apart from my country, where technology is something that has not been seen for years, in countries like the United States, Canada, Germany, China, Russia, and all the countries that are recognized as being among the leaders in this field of technology, there are still cases where the killer or criminal remains unidentified and never faces the consequences of their actions.

And with this technology, they can use the smallest clue to catch the killer—a technology that, if you tell people in a third-world country about it, they will say you are crazy. A technology that, if they do not understand the logic behind it, they will think it is magic, while in reality, it is just science and nothing more. The name of this scientific magic is DNA Phenotyping.

Biological and Genetic Foundations of DNA Phenotyping

To understand exactly what DNA phenotyping does, first we need to understand an important difference from old DNA methods. For years, DNA has been used in criminal investigations through something called **STR**. You have probably heard the term “DNA fingerprinting.” This method basically examines specific parts of DNA and then says: “Well, does this DNA match this person’s DNA or not?”

STR is very good for finding a connection between a sample from a crime scene and a known suspect. For example, if a person’s DNA has already been registered in a criminal database, the police can compare the crime scene sample with it. But there is a relatively big problem: if the killer’s DNA is not in the database, STR cannot do much special anymore. It basically says: “Well... we have DNA, but who does it belong to? You find out yourself!” and from it, it cannot be understood what the person looks like.

This is where **DNA Phenotyping** enters the story.

DNA phenotyping, instead of only looking for similar DNA in a database, tries to obtain information about a person’s possible appearance from DNA. For this, scientists look at small changes in DNA called **SNP**.

SNP can be imagined very simply like this: in a specific part of DNA, one person’s genetic letter may be different from another person’s. This small difference can sometimes be related to characteristics such as eye, hair, or skin color.

Scientists examine many of these small differences and, using statistical models, calculate the probability of certain physical characteristics. Of course, DNA is not going to print a

picture of the killer's face and say: "Here you go, this is the killer!" The story is a little more complicated than that.

One of the well-known systems for predicting eye, hair, and skin color is **HirisPlex-S**. This system uses several genetic markers to estimate the probability of certain physical characteristics based on DNA.

Genetics of Eye and Hair Color

Eye color does not depend on only one gene and several genes play a role in it. But one of the most important parts is located on **chromosome 15** and two important genes called **OCA2** and **HERC2** are located in this region.

The OCA2 gene plays a role in the production and regulation of pigment. Alongside it, a genetic change in HERC2 can affect the activity of OCA2. In simple words, we can say that this part of DNA acts like a control switch; of course, a switch that instead of turning a lamp on and off, affects the amount of pigment in the eye.

If OCA2 activity is higher, there is usually more pigment in the iris and the probability of brown eyes becomes higher. If its activity is lower, the amount of pigment becomes lower and the probability of lighter eyes, such as blue, increases.

Of course, the story is not this simple. Colors such as green and hazel are the result of a combination of several genetic factors and genes such as **TYR**, **TYRP1** and **SLC24A4** can also play a role in this matter.

Hair color has its own story. Our hair is mainly affected by two types of pigment:

- **Eumelanin** → mostly related to brown and black colors
- **Pheomelanin** → mostly related to red and yellow colors

One of the well-known genes in this story is **MC1R**. This gene plays a role in regulating the type of pigment that the body produces.

Some changes in MC1R can cause the body to produce more pheomelanin and as a result, the probability of red hair becomes higher. But here too, we should not think that DNA is like a vending machine where we put in an SNP and a red-haired person comes out from the other side! Several genes and different factors play a role in hair color.

Genes such as **KITLG** and **SLC45A2** can also affect pigments and hair color and can be examined in genetic prediction models.

Genetics of Skin Color

Skin color is even more complicated than this. Human skin color is related to different genetic factors and also to the long-term adaptations of humans to the amount of sunlight.

Systems such as HIrisPlex-S can estimate skin color in several general groups by examining a set of genetic markers; for example, very light, light, intermediate, dark and very dark.

Two important genes in this matter are **SLC24A5** and **SLC45A2**. Specific changes in these genes can affect how pigment is produced and stored in skin cells.

For example, some changes in SLC24A5 are associated with pigmentation differences between different human populations. SLC45A2 also plays a role in the process related to melanin production.

By putting together the information related to these genes and several other markers, scientists can provide a statistical estimate of a person's skin color.

Of course, one important point should not be forgotten: this method is not supposed to say from DNA "this person looks exactly like this, even their clothes are probably this color!" DNA only predicts certain characteristics with a specific probability. In other words, it is more like a **very carefully calculated scientific guess**, not a hidden camera inside DNA.

In the end, the attractiveness of DNA Phenotyping is exactly here: something that at first glance may look like magic is actually a combination of genetics, biology and statistics. DNA itself does not point at the killer with a finger; but it can provide clues about the possible appearance of the person to investigators.

Analytical Core: When the Computer Starts Guessing What a Person Looks Like from DNA!

So far, we understood that DNA can give us information about eye, hair, and skin color. But the story becomes more interesting when scientists want to guess things about the **shape of the face** from DNA as well.

Of course, here the matter is no longer that the computer looks at the DNA and says: "Aha! Here's the person's face, please take it!" Unfortunately, we still don't have such magic.

The shape of the human face depends on so many things; for example, the shape of the jaw, the size of the nose, the distance between the eyes, the shape of the bones of the face and a whole lot of other things. On top of that, thousands of different genes can also affect these features. Meaning the human face is kind of a group project that almost thousands of genes are involved in, and each one is playing its own tune.

That's why scientists use **machine learning**; meaning they give the computer a lot of information so that it gradually finds the patterns by itself.

How does the computer try to build the face?

First, there needs to be a large dataset. In different researches, people's genetic information is collected alongside information related to their faces. For example, 3D scans of people's faces can be used so that the exact shape of their faces is recorded.

Then the computer starts looking for the relationship between genetics and face shape.

One of the methods used to simplify this huge amount of information is **PCA**. Its name sounds very scary, but its main idea is not that strange.

Imagine a 3D face has a huge amount of information and the computer almost freezes when seeing it. PCA helps turn this complex information into a few more important and manageable parts.

Then different machine learning algorithms try to understand:

"If we see this genetic pattern, which facial feature will probably change as well?"

For example, some genes like **EDAR** and **PAX3** can be related to some features of facial development. Therefore, if the statistical model finds a specific relationship between some genetic changes and the shape of a part of the face, it can put that information into building the 3D model of the face.

In the end, the system can build a **probable image of the face**.

Meaning raw DNA, which to the human eye just looks like a bunch of meaningless letters and codes, turns into something that an investigator can look at.

Of course, I emphasize again: this is not a real photo of the killer. The computer is not going to pull the person out of the DNA in 4K quality and say: "Here you go, this is the same guy, and he's even going to trim his beard tomorrow!" This is only a prediction based on genetic information.

The Big Problem: DNA Doesn't Know Everything!

Here there is an important problem.

DNA has a lot of information, but it **doesn't know everything**.

For example, DNA cannot understand how much food someone has eaten over the last 30 years, how many times their nose has been broken, or whether they shaved their beard this morning or not.

Age is one of the important factors. The face of a child, a young person and a 50-year-old person is naturally not the same.

Weight also affects the shape of the face. Nutrition, living conditions and even injuries such as breaking the nose can change a person's appearance.

Now imagine that the computer only has the DNA of a 50-year-old person and builds a probable face based on genetics. The result may be similar in some features to a younger version of that person.

Meaning the computer might say:

"According to the DNA, this is what he looks like."

And the person answers:

"Bro, I'm 50 years old now, that photo is from high school!"

Things like beard, mustache, hairstyle, tattoos, scars and cosmetic surgeries are another problem. DNA is not going to understand that the person decided to get their nose operated on or grow a three-meter beard for themselves.

Changes related to the environment and lifestyle can also affect appearance. Even some **epigenetic** changes, meaning changes that affect how genes function without changing the original DNA sequence itself, can play a role in some characteristics of the body.

So we must always remember that the result of DNA phenotyping is a **probabilistic prediction**, not an identity-card photo.

The Scariest Part of the Story: Ethics and Privacy

Now imagine the police can obtain a lot of information about a person's appearance and genetic characteristics from DNA.

Well, the important question is:

Does everything we can understand scientifically, also mean that we are ethically allowed to understand it?

This is where the story gets complicated.

The human genome does not only have information about eye and hair color. Genetic information can also have clues about family ancestry and some biological and medical characteristics.

Therefore, if an organization examines all of a person's genetic information, the issue is no longer just finding out eye color. Very private information is involved.

In simple words, DNA is not like an ordinary file where if someone sees it, you say "well, you saw it, so what."

DNA is more like a **very personal and extremely large file** that may contain information about the person themselves and even their family.

The Next Problem: The Computer Can Make Mistakes Too!

We usually think when we hear the name "artificial intelligence" or "algorithm," that the computer is a genius creature that never makes mistakes.

No.

The computer can also make mistakes very easily; it just makes its mistake with more confidence!

One of the important problems in this field is **algorithmic bias**.

If the information used to train a model is mostly related to a particular group of people, the model may have lower accuracy for other groups.

For example, if most of the samples available in a database are from people of European ancestry, the model may perform less accurately when predicting the characteristics of people from other populations.

This issue is important in criminal investigations, because a wrong prediction is not just a wrong number on a computer screen.

It may cause the police to focus more on a particular person or even a particular group, while that person may not be a criminal at all.

And this is where a small computer mistake can turn into a very big human problem.

So Can You Consider Someone a Criminal Just with Their DNA?

Here the answer must be very clear:

No.

DNA phenotyping is more suitable for creating a **clue**, not for saying:

“Well guys, DNA said it’s this person, close the case!”

For example, if there are no clues in a case, DNA information may help investigators narrow down the search range.

It may also create new information in old cases that have remained without results for years.

But the image created by the computer is a probability, not definite evidence of a person’s identity.

In fact, the best way to look at this technology is to consider it a **highly advanced investigative assistant**.

It is a good assistant, but we still shouldn’t give the entire case to it and say:

“You solve it, we’re going to drink tea.”

Because in the end, to accuse a human being of a crime, we need real and verifiable evidence, not just a probable face that an algorithm built from DNA.

And perhaps this is the most interesting part of the story: **DNA can give a clue, the computer can make a guess, but neither of them replaces real investigation.**

Conclusion

If we want to sum up this whole story in a few sentences, DNA phenotyping, or **FDP**, is one of the most interesting things that has entered forensic science in recent years. Something that once looked more like a scientific and theoretical idea has now, with the help of DNA sequencing, statistics and machine learning, become a real tool.

Until quite recently, DNA was mostly used to say: “Does this sample match this person’s DNA or not?” But now the story has become a little stranger. Scientists can obtain information about a person’s possible physical characteristics from a very small amount of DNA traces. Systems like **HirisPlex-S** can give clues about eye,

hair and skin color, and newer methods are also trying to extract more information about the shape of the face.

This means that a case that has had no proper clue for years might come back to life. Of course, we shouldn't get too excited here and think that from tomorrow the police will find a strand of hair, put it into a machine, and five minutes later receive a 4K picture of the killer!

The reality is a little more complicated.

Despite all its advances, this technology still has many limitations. DNA cannot understand what food someone has eaten throughout their life, how many times their nose has been broken, how much weight they have gained, what hairstyle they have, or for example, that they decided to grow a beard and turn their face into a historical character!

On the other hand, algorithms are not magical creatures either. If the information they are trained with is incomplete or unbalanced, the result can also be wrong. Especially if samples related to some populations are much more common than others, the prediction for other groups may have lower accuracy.

And this mistake in forensic science is no joke. If an algorithm makes a mistake, the issue is not just that the computer gave a wrong answer. It may cause an innocent person to be suspected more or cause the police investigation to go in the wrong direction. That is why the issue of genetic privacy and the possibility of algorithmic bias must be taken completely seriously.

In the end, in my opinion, the best way to look at DNA phenotyping is not to consider it a **killer-finding machine**. This technology is more like a very cool detective that has a lot of information, but can still make mistakes sometimes and shouldn't be given the entire case!

FDP can provide clues, narrow down the direction of an investigation and even create hope again in some old cases; but it should not determine on its own who is guilty.

In the end, no matter how much technology advances, there is still one thing we cannot hand over to a computer and go on with our lives: **judging a human being**.

Science can tell us what is likely to be true. An algorithm can examine millions of pieces of data. DNA can give us strange clues. But what conclusion we draw from this information still requires human investigation, examination of evidence, and proper reason and judgment.

Perhaps the most fascinating part of this whole story is exactly this: we have reached a point where a very small piece of DNA can tell us things about its owner that would

have been almost unimaginable a few decades ago. But this greater power also brings greater responsibility.

So no, DNA is not magic, artificial intelligence is not detective Sherlock Holmes, and the computer is not going to hand us the killer with their home address tomorrow morning based on a single strand of hair!

But if this technology is used correctly and responsibly, it can be one of the very powerful tools of forensic science; a tool that instead of replacing humans, helps humans see things that they previously could not see at all.

References

Kayser, M. (2015). Forensic DNA phenotyping: Predicting human appearance from crime scene DNA for investigative purposes. *Forensic Science International: Genetics*, 18, 33–48.

Parabon NanoLabs. (2025). *Snapshot DNA Phenotyping: System Overview and Case Studies for Law Enforcement*. Forensic Technology Technical Report.

Schneider, P. M., Prainsack, B., & Kayser, M. (2019). The use of forensic DNA phenotyping in criminal investigations and courts. *Deutsches Ärzteblatt International*, 116(51–52), 873–880.

Walsh, S., et al. (2018). Global skin colour prediction from DNA: Introducing the HirisPlex-S system. *Forensic Science International: Genetics*, 34, 214–223.

Declaration of Generative AI and AI-Assisted Technologies

Statement of Authorship:

The core ideas, arguments, structure, and conclusions of this paper were independently developed and authored by **Mohammad Mostaan Abedi**.

Use of AI Tools:

ChatGPT and other AI tools were used only to assist with:

1. Translating and rephrasing the author's original Persian/Dari text into academic English.
2. Correcting grammar and syntax.
3. Verifying technical terminology, including genes and machine-learning concepts.

The AI tools did not generate the paper's original ideas, arguments, or conclusions. The author retains full responsibility for the final content.