Why Not Use Computer Algorithms for Polygraph Data Analysis?

Article · January 2024

1 author:

Raymond I Nelson

WHY NOT USE COMPUTER ALGORITHMS FOR POLYGRAPH DATA ANALYSIS?

Raymond Nelson

With advancements in technology, automated computer algorithms have become increasingly thought of as a useful adjunct to traditional methods of manually scoring Psychophysiological Detection of Deception (PDD) test data. While computer algorithms offer potential benefits such as consistency and efficiency, there are several reasons why field practitioners may wish to avoid their use. These arguments may also interest program managers responsible for developing and implementing polygraph field practice policies, supervisors and quality assurance personnel who oversee the execution of polygraph methods, along with subject matter experts involved in the review of polygraph examinations and polygraph programs.

The reasons for refraining from using automated computer algorithms are several, and include ethical concerns when making decisions about individuals, and concerns about potential over-reliance on technology, statistical methods, and science in general. Some expressed concerns involve matters of reliability, validity, and the potential for misclassification errors. Additionally, the use of automated PDD scoring algorithms has been cautioned against because it can flatten the expert development curve and lead to the devaluation of human expertise.

Another concern is that the widespread use of automated data analysis algorithms would result in younger professionals not learning, and departing from, the traditional methods. Computer algorithms have also been discussed with some degree of mistrust because they may approach the analytic task differently from human scorers. Furthermore, field examiners may feel compelled to report the result from a computer algorithm if it differs from their own conclusion, potentially undermining their professional judgment. Widespread reliance on automated computer algorithms might prevent PDD field examiners from making use of new ideas and idiosyncratic solutions. And finally, there is concern that field examiners, who are not involved in the development and validation of these algorithms, may be unable to answer questions about their functioning, development, and validation in a courtroom or legal proceeding.

Ethical Concerns

Computer algorithms use complex mathematics, statistics, and procedures that may be opaque, making it difficult for PDD examiners and other professionals to understand how decisions are made. This raises ethical issues related to transparency and accountability when making decisions that impact the lives of other individuals. Ethical discussions should consider the level of knowledge that practitioners might have about these algorithms, including their development, validation, use, functionality, and interpretation. Ethical discussions should also be informed about the topic of ethics itself.

Ethics involves understanding the difference between good and bad actions. Various ethical paradigms have been suggested. Utilitarian ethics considers actions ethical if they result in the greatest overall benefit relative to the harm caused. This framework underpins much of our legal systems but has also been used to justify morally wrong practices such as slavery and discrimination.

Deontological ethics argues that it is unethical to use any individual as a means to an end if it causes them harm. Each person is an end unto themselves, with ethical obligations owed to every individual. Medical and helping professions are founded on this principle, encapsulated by the phrase “first, do no harm.” While promoting individual rights, this approach can be impractical if taken to extremes, potentially hindering essential social structures due to the risk of harming an innocent person.

Virtue ethics focuses on developing virtues such as trustworthiness, loyalty, dependability, compassion, humility, fairness, prudence, and honesty. An individual’s ethical standing is determined by their adherence to ethical virtues. However, this paradigm can be somewhat circular, and can fall short in addressing real-world problems, particularly with individuals who possess a mix of ethical and unethical traits.

Professional ethics are often declarative, requiring adherence to established policies and guidelines that provide clear standards for conduct. This ensures accountability and consistency but offers limited guidance on discussion about whether the rules themselves are ethical. This raises critical questions: Is there such a thing as a bad rule? How do we know if a rule is bad? And, what should be done?

In PDD data analysis, effective practices must balance technological benefits with potential harms. In a very practical sense, it is important to consider what potential harm could come from the use of automated scoring algorithms. Crafting ethical guidelines for using advanced technologies will likely require a dialectical approach that leverages the strengths and acknowledges the limitations of various ethical paradigms.

One important question is whether using a computer-based classifier for deception and truth-telling represents a shift of ethical responsibility. Is it ethical for a machine to make important decisions about a person? Another consideration is whether it is ethical to forgo use of the most effective classification methods. A more nuanced matter involves the degree of fluency and competency field practitioners should have in statistical classification and decision making.

Addressing some of these ethical questions is straightforward. Important decisions about people should be made by professionals, not machines. Algorithms used for important decisions should be documented and accountable, with details available for review by stakeholders. Oversight and scrutiny by other subject matter experts are essential to ethical professional practice. However, it is unrealistic to expect PDD field practitioners to understand and replicate all mathematical and statistical procedures or inspect computer code for errors. Instead, field examiners should focus on effectively understanding and interpreting algorithm outputs.

Validity, Reliability, and Potential Misclassification

One of the arguments against the use of automated computer algorithms in PDD data analysis is the potential for misclassification. More broadly, this concern involves the validity and reliability of computer algorithms. In addressing these concerns, it is important to correctly understand the conceptual terminology. Discussions about reliability in a scientific context refer to whether the results from a scientific test or experiment can be expected to be observed again if the test or experiment, or the analysis, were repeated. The notion of reliability is sometimes easily conflated with validity. Reliability pertains to whether a test consistently produces the same results under the same conditions. Validity, on the other hand, concerns whether the test accurately measures what it is intended to measure – whether the output is an accurate description of reality. Discussions about polygraph validity, for practical purposes, pertain to the accuracy of output classifications of deception or truth telling.

There is no doubt that a test or data analysis algorithm that is incapable of misclassification error would be highly desirable. For this to occur, it would have to be immune to both random variation and all forms of behavior on the part of both the examinee and examiner. In reality, there is no form of scientific test or analysis method that is without some potential for error – due to either human factors or uncontrolled/random variation.

Algorithms, like all scientific tests, cannot reasonably be expected to be infallible. They can produce both false positive and false negative outcomes, which can have serious consequences, particularly in legal and employment contexts. It is important to note that scientific tests, and the algorithms used to analyze test data, are inherently probabilistic, intended to quantify phenomena that cannot be subject to direct physical measurement. While the risk of misclassification is an important concern, developers of test data analysis methods, whether manual or automated, strive to reduce or minimize the potential for testing error. This begins with the extraction of relevant information from the test data and the isolation and rejection of unwanted noise. Ultimately, questions about computer algorithm validity will focus on whether the resulting classification of deception and truth telling, along with the potential for misclassification error, achieve acceptable and usable statistical effect sizes – and whether validation studies show these to equal or exceed the effect sizes from manual TDA methods.

The validation process is similar for both manual and automated TDA algorithms. It traditionally involves developing a classifier on one dataset, followed by the inspection of effect sizes on another dataset. The goal of validation is not perfection or infallibility, but to estimate the range of expected effect sizes on data other than the training data and to avoid the tendency for over-fitting and overestimation that can occur under naive conditions when development and validation are based on the same sample data.

The goal of validation is to calculate realistic estimates of expected effect sizes that can be observed in field practice. Statistical computing platforms have contributed to important changes in validation techniques. For example, techniques now exist to calculate usable ranges for effect size estimates using statistical validation and cross-validation methods using a single sample. Also, computer simulation techniques can provide insight into complex systems that might be difficult to analyze using traditional methods. Interestingly, automated PDD data analysis algorithms have performed as well as or better than most expert human scorers in published studies.

One of the differences between manual scoring methods and automated computer algorithms is that manual scoring procedures require rigorous procedural training, in addition to ongoing supervision and quality control, in order to achieve adequate reliability. In contrast, computer algorithms may require increased conceptual knowledge, but can execute complex procedures with automated reliability. Another difference is that whereas manual scoring procedures are expected to become less reliable with increased complexity, computer algorithms are less vulnerable to fatigue and social influences, and maintain their expected validity and reliability regardless of the complexity of the methods.

Importantly, the reliability of any test or analysis method sets the upper limit for its validity. In other words, a test cannot be more valid than it is reliable. With this in mind, and assuming all other factors are equal, human professionals using manual scoring methods will ultimately be less reliable than automated solutions for any correctly structured complex procedure. This means that the potential validity of any manual scoring method is ultimately lower than that of a similarly designed automated process. This is based on the requirement that the analysis process is known and not heavily reliant on unstructured judgment.

On the other hand, if valid PDD outcomes are highly dependent on clinical procedures that are so complex that they cannot be fully structured and automated, then all attempts to develop and validate automated PDD algorithms are likely to be unsuccessful. Similarly, efforts to quickly and easily develop expert skills and achieve reliable outcomes among a large group of diverse professionals would also likely be unsuccessful for any unstructured highly complex clinical method for TDA.

Over-Reliance on Technology, Statistics, and Science

Another argument against the use of automated computer algorithms for the analysis of PDD test data is that it results in an overreliance on technology, statistics, and science. Some believe this overreliance may undermine the role of human expertise, potentially causing PDD field practitioners to neglect the development of their own skills and subjective intuition, and, become dependent on automated systems. Even more to the point, if statistical models do not capture the complexities and subtleties of human physiological responses, the resulting probabilistic inferences about deception and truth-telling might not account for individual differences and context-specific factors. In this case, computer algorithms can be expected to fail to achieve desired effect sizes. Importantly, effect sizes for some automated PDD scoring algorithms have equaled, or exceeded, those of most human experts in published studies.

While it is important to avoid over-reliance on technology, integrating human expertise with automated systems can enhance both the reliability and accuracy of polygraph examinations. Algorithms can process data quickly and consistently, providing preliminary analyses that human examiners can then review and interpret. This ensures that the strengths of both technology and human judgment are effectively utilized and highlights the ethical role of human experts when making decisions that affect individuals.

Defining the boundary of "over-reliance" on science, technology, and statistics can be challenging. However, the key here is to ensure that computer algorithm outputs serve as an aid to human judgment rather than a replacement. Integrating automated systems with human expertise can lead to a more robust and useful analysis. For instance, automated algorithms can flag inconsistencies or patterns that might be overlooked by human examiners due to cognitive biases or limitations in processing high-dimensional data. Conversely, human examiners can bring contextual knowledge, intuition, and ethical considerations into the decision-making process, which algorithms may not fully grasp.

Flattening of the Professional Development Curve and Devaluation of Expertise

The use of automated PDD scoring algorithms can flatten the expert development curve, leading to the devaluation of expertise and the loss of human judgment. Expert PDD field practitioners – with years of training and experience – can sometimes interpret subtleties and details in physiological data that algorithms might overlook. Automated algorithms may create the impression that "anyone can do it," undermining the specialized skills and deep knowledge required for accurate polygraph analysis. There is also the concern that the roles or jobs of human experts might eventually be completely replaced by robots or autonomous systems. This is, however, unlikely to occur.

Flattening the expert development curve may ultimately be an advantage, not a deficit, as could ensure that a larger number of field practitioners can deliver analytic results with high reliability and high validity. Automation of complex and routine tasks would permit human experts to focus on more complex and nuanced aspects of their work that are outside the scope and capabilities of automated algorithms. Integration of human expertise and technological advancements can lead to a more efficient workflow in which the strengths of both can be leveraged to improve the practical value of PDD test results.

Younger Professionals Might Not Learn Traditional Methods

As technology continues to evolve and integrate into various fields, younger professionals may increasingly rely on new methods and tools, potentially at the expense of developing traditional skills. This trend is noticeable in many fields, and particularly where automated systems and algorithms are becoming more prevalent. While technological advancements can bring numerous benefits, such as increased efficiency and consistency, there is concern among some experienced PDD field practitioners that younger professionals might not learn the “old ways” of manual scoring and traditional/visual PDD data analysis.

The knowledge and skills developed through traditional TDA methods provide a deep understanding of the fundamental principles and of PDD test data. This foundational expertise can enhance critical thinking, problem-solving, and adaptability. However, with a heavy reliance on automated tools, younger professionals might miss out on this useful aspect of their training. They may become proficient in using advanced software and algorithms but lack the direct experience and intuition that comes from manual scoring of test data.

Moreover, traditional methods can involve a level of craftsmanship and professional insight that is difficult to replicate with technology. These methods can foster a more comprehensive understanding of the data and its interpretation. As the PDD profession evolves, it may be useful to find a balance that allows younger professionals to benefit from technological advancements while still acquiring the valuable skills and insights that come from traditional approaches. This would require training programs to incorporate a blend of old and new methods, ensuring that younger professionals are well-rounded in their expertise. By preserving the teaching of traditional techniques alongside modern technologies, the polygraph profession can maintain a high standard of practice and prepare the next generation of professionals to be both technologically savvy and deeply knowledgeable about their field.

Algorithms Analyze Data in Ways that Differ from Human Scorers

One of the fundamental differences between algorithms and human scorers is in their basic approach to data analysis. Algorithms are designed to process data systematically and consistently, following predefined mathematical and statistical models. This enables algorithms to detect patterns and correlations that might not be immediately apparent to human scorers using unstructured or semi-structured visual methods. In contrast, human scorers rely on their training, experience, and intuition to interpret physiological responses, which can introduce subjectivity and variability into the scoring process. This subjectivity can be influenced by cognitive biases, fatigue, social factors, and other human factors such as attachment to or confusion with particular hypotheses or practices, leading to inconsistencies in the interpretation of the same data.

Importantly, algorithms can execute complex calculations rapidly and without error, providing objective and reproducible results. They are particularly adept at handling high-dimensional data and can integrate various physiological signals to produce a comprehensive assessment. Human scorers, on the other hand, might prioritize certain signals over others based on their judgment or bias, potentially missing or ignoring useful indicators of deception or truth telling. While human expertise is invaluable for contextual interpretation and ethical considerations, the consistency and objectivity offered by algorithms can help to ensure that the data is analyzed uniformly, reducing the likelihood of error and bias.

Although it might be tempting to expect that human experts and automated algorithms would approach data analysis tasks similarly, it is ultimately naive to hold such expectations without consideration for both examiner training and algorithm design. Field PDD practitioners can be trained to use some of the same statistical methodologies that computer algorithms employ. Additionally, computer algorithms can be designed to use only the methodologies that are supported by empirical evidence and commonly employed by human experts. However, there are inevitable limits to the computational burden that human experts might enjoy or endure, and there are distinct advantages to the unconstrained use of powerful statistical computing platforms.

Ultimately, it is unrealistic to expect human experts and computers to approach data analysis tasks in exactly the same way. Differences in scoring methods highlight the potential for a hybrid approach, where algorithms provide a preliminary, objective analysis that human examiners can review and contextualize, combining the strengths of both methodologies. This integration ensures a more balanced and comprehensive analysis, leveraging the speed and consistency of algorithms along with the highly developed expert judgment and contextual understanding of human experts.

Examiners May Feel Compelled to Report the Computer Algorithm Result

There is some potential, when using automated TDA algorithms, that PDD field examiners may begin to feel compelled to report algorithm-generated results, even when they differ from their own conclusions. This dilemma is particularly challenging in high-stakes contexts such as legal or employment settings, where human professionals are ultimately responsible for the correct execution and reporting of the test and its analytic results.

When an examiner's interpretation of PDD test data differs from an automated scoring algorithm, the examiner must decide how to reconcile the difference, and which result to use as a basis for their professional opinion. Algorithms process data consistently using objective criteria, enhancing reliability. However, human experts bring experience and contextual understanding to their analysis, sometimes considering details and factors that algorithms might overlook.

Deference to algorithmic results can stem from perceptions that use of a computer algorithm requires that the algorithm output should automatically become a basis for the examiner's professional opinion – or that algorithm output should always coincide with the results from manual TDA procedures. Deference to algorithm results could also result from a perception that algorithms are unbiased and free from human error, lending their results an unwarranted aura of infallibility. Organizational policies and legal frameworks may favor standardized, reproducible methods, leading examiners to feel that their professional judgment is undervalued or that they are vulnerable to criticism if their conclusions do not align with algorithmic outputs.

Conversely, some organizations may prohibit the use of automated TDA algorithms to avoid perceived inconsistencies or vulnerabilities. This can introduce other issues. For example, in a courtroom, an examiner might first encounter the algorithmic result during a legal proceeding, with little time to reconcile it with their own findings. Highly effective organizational policies might require the integration of manual and automated results prior to the submission of results for a legal proceeding.

Automated Data Analysis Algorithms Limit the Ability to Make Use of New Ideas

Automated data analysis algorithms, while highly efficient and consistent, have been subject to some criticism because they limit the ability of PDD field practitioners to incorporate new ideas and innovative approaches. One of the key issues with automated systems is their dependence on existing datasets and established methods. While this can enhance accuracy and reliability, it also means that these systems are less likely to easily integrate new and untested approaches that might offer valuable insights. Human practitioners, in contrast, have the ability to think outside the box and apply creative or idiosyncratic solutions to complex problems. This kind of innovative thinking is difficult to encode into an algorithm.

In practice, however, constraining data analysis to standardized practices and evidence-based methods may represent an advantage, rather than a disadvantage. The use of unstandardized and untested methods in field practice poses several hazards that can undermine the reliability and validity of outcomes. Without thorough testing and validation, it is impossible to know how new ideas and idiosyncratic methods will perform or whether they will produce accurate results. Moreover, reliance on unstandardized and untested methods can compromise the accountability and professionalism of individual practitioners.

In the absence of standardized procedures, it becomes challenging to ensure that all practitioners are adhering to the same high standards of practice. Lack of standardization and consistency can create a perception of unreliability and diminish the credibility of the entire profession. The adoption of standardized and validated methods is essential for maintaining the integrity and reliability of field practices. Although not all professional practices can be subject to automation, automation can be thought of as the ultimate form of standardization, and should be considered wherever possible.

Field Practitioners Cannot Answer Questions about Algorithms

It has been suggested that PDD field practitioners should not use automated scoring algorithms because they cannot adequately answer questions about their design, development, and validity, and cannot adequately describe how a result was achieved. It has been suggested that field practitioners, who may lack a deep understanding of the technical aspects of an algorithm, may be unable to defend their findings in high-stakes environments such as a legal proceeding.

This concern seems to overlook the fact that any high-stakes legal proceeding will involve the services of subject matter experts who may offer differing opinions based manual scoring methods that are more subjective and therefore less objectively describable than result obtained using automated algorithms. It also overlooks that computer scoring algorithms are not likely to be utilized without publication of their validity – which should be available to field practitioners and other stakeholders. More importantly, it seems to be based on a misunderstanding of the roles and responsibilities of field practitioners versus algorithm developers.

Field examiners are not expected to be algorithm developers or possess the technical expertise to explain the intricate details of algorithm design and validation. Their primary responsibility is to use these tools correctly and interpret the outputs in a meaningful and contextually relevant manner. Just as a radiologist is not required to understand the engineering behind MRI machines, PDD examiners do not need to master the technicalities of algorithm development.

Questions about algorithm design, development and validation should be asked of algorithm developers. These questions may involve details pertaining to feature extraction, artifact rejection, statistical quality control, numerical transformation and data reduction, statistical and structural models, procedural rules for parsing categorical result from probabilistic information, expected effect sizes (sensitivity, specificity, error rates, and other accuracy metrics), and other aspects of algorithm development. Interestingly, PDD field examiners are not required to be conversant with the answers to these questions as they pertain to manual scoring methods – but are required to use manual scoring methods correctly.

Conclusion

The debate over using automated scoring algorithms for polygraph data parallels broader discussions about technology integration in various fields, such as digital calculators in mathematics, autopilot systems in aviation, adaptive safety systems in cars, and algorithms in medical procedures like LASIK. These comparisons illustrate the potential benefits of technology integration.

Critics of automated polygraph data analysis highlight issues like ethical concerns, misclassification risks, over-reliance on technology, and the devaluation of human expertise. These challenges can be mitigated through proper training, integration of human expertise, and effective field practice policies.

Polygraph testing sits at the intersection of scientific inquiry and practical application. For the polygraph profession's credibility and efficacy, it is essential to integrate technological advancements within a scientifically rigorous framework. Combining technology with scientific and practical expertise will help in maintaining professional integrity and securing legislative, scientific, and community support.

A collaborative approach that includes both human and algorithmic analyses will enhance polygraph examinations. Documenting reasoning and factors leading to conclusions builds trust and professionalism. Automated algorithms can improve PDD testing by efficiently processing data and recognizing patterns, enabling human examiners to focus on interpretation and context.

Field examiners should be knowledgeable about integrating algorithmic tools, understanding their outputs, strengths, and limitations. Examiners should be fluent in probabilistic concepts like alphas, p-values, confidence intervals, sensitivity, specificity, and conditional metrics such as posterior probabilities and odds ratios, all of which relate to the discussion of test outcomes. This fluency allows examiners to contextualize algorithmic outputs, ensuring professional judgment remains central to the process.

Organizations should implement policies to handle discrepancies between human and algorithmic results, establishing protocols for reviewing and reconciling differences and escalating cases when necessary. This maintains transparency and accountability, emphasizing the ultimate responsibility of human experts for analysis and results.

In summary, integrating computer algorithms into polygraph data analysis enhances efficiency and accuracy. By balancing human expertise with technological advancements, the polygraph profession can ensure high reliability, accuracy, and integrity in its practices.