Using AI Innovations to Improve Accuracy and Reliability in Medical Testing

Yuming Xiao *

Shenzhen Foreign Language School, Shenzhen, China

* Corresponding Author Email: 2180804319@qq.com

Abstract. While medical testing serves as a vital pillar of modern clinical diagnostics, conventional methodologies—most notably fixed-threshold models and frequentist statistics—often struggle to account for data noise and inherent biological variability. These rigid, "one-size-fits-all" standards frequently lead to diagnostic misjudgments and fail to provide a quantified measure of uncertainty in complex scenarios. This paper investigates the role of artificial intelligence (AI) in bridging these gaps to refine both diagnostic precision and reliability. We first examine how Bayesian probabilistic frameworks, rooted in Bayes'Theorem, can shift diagnostics from binary outcomes to personalized probability assessments by accounting for uncertainty. The discussion then moves to the practical application of Deep Learning, particularly Convolutional Neural Networks (CNNs), in automating the detection of pathological features in screenings for conditions such as lung cancer and diabetic retinopathy. Additionally, we explore how transfer learning mitigates the challenge of limited datasets in rare disease research, while sequence models offer insights into longitudinal patient history. By synthesizing these advancements, this paper argues that the integration of AI facilitates a necessary transition toward more robust, individualized diagnostic frameworks within precision medicine.

Keywords: Clinical Diagnostics; Artificial Intelligence; Bayesian Inference; CNNs; Personalized Medicine; Diagnostic Reliability; Transfer Learning; Medical Imaging Analysis.

1. Introduction

Medical testing is the foundation of modern healthcare, which supports disease screening, early diagnosis, and personalized treatment. However, real-world test data often contains noise, such as instrument errors, sample contamination, and individual biological differences. These uncertain factors make traditional testing methods prone to misjudgments. For a long time, the threshold method and frequentist statistical analysis have been widely used in clinical practice. They are simple and low-cost, but cannot meet the needs of precision medicine. With the development of artificial intelligence, Bayesian probabilistic models, deep learning, and sequence models provide new solutions to the limitations of traditional methods. This essay will explain the shortcomings of traditional medical testing methods and discuss how AI technologies improve the accuracy and reliability of medical testing.

2. Traditional Medical Testing Methods

Two traditional methods are commonly used in clinical testing, but both have obvious and unavoidable defects.

The first one is the threshold (cutoff) method. It sets a fixed standard for biological indicators based on population data and clinical guidelines. For example, according to the American Diabetes Association (2023), a fasting blood glucose level ≥ 7.0 mmol/L is used for diabetes screening, and a CEA level > 5.0 ng/mL may prompt further cancer evaluation. This method is cheap and easy to use, so it is suitable for large-scale screening and low-resource areas. However, it uses a rigid "one-size-fits-all" rule, completely ignoring individual differences. It only provides binary results and cannot quantify the uncertainty caused by data noise, which greatly reduces its reliability in complex cases.

The second one is frequentist statistical analysis. It evaluates test performance through population-level indicators such as sensitivity, specificity, and overall accuracy. For example, large clinical trials confirm that a COVID-19 PCR test has 98% accuracy (Wang et al., 2020). This method is often used to verify new testing tools, but it has a fatal flaw: population statistics cannot be directly converted into an individual‘s disease probability. In rare disease testing, even a highly accurate test will produce many false positives. Over-reliance on this method will overestimate personal risk and lead to unnecessary medical examinations and interventions.

3. AI Innovation 1: Bayesian Probabilistic Models

Bayesian probabilistic models are core AI methods to solve data uncertainty. The foundation of these models is Bayes' Theorem, which provides a mathematical framework for updating the probability of a hypothesis based on prior knowledge and observed evidence. Bayes' Theorem is expressed as:

$$P(D|T) = [P(T|D) \times P(D)] / P(T) \tag{1}$$

Where:

- P(D) is the prior probability, representing the estimated probability of disease before testing, often based on disease prevalence or individual risk factors;
- P(T|D) is the likelihood, typically the test's sensitivity (true positive rate);
- P(T) is the marginal probability of a positive test result;
- P(D|T) is the posterior probability, the updated probability of disease after considering the test result.

Unlike traditional methods, Bayesian models do not output simple positive or negative results, but provide a specific disease probability with a reliable range. For example, consider a rare disease with a prevalence rate of 1% in the population and a test with 99% sensitivity and 99% specificity. A naive interpretation might assume that a positive result indicates a 99% chance of disease. However, applying Bayes' Theorem yields:

$$P(D|+) = (0.99 \times 0.01) / (0.99 \times 0.01 + 0.01 \times 0.99) = 0.50 \tag{2}$$

That is, the true probability of disease after a positive test is only 50%. This counterintuitive result was first highlighted by Casscells et al. (1978) in a seminal study on physician misinterpretation of diagnostic tests, and further emphasized by Manrai et al. (2016) as a critical cause of overdiagnosis in rare disease screening. Bayesian models thus help clinicians avoid unnecessary medical interventions by providing accurate, personalized probability estimates. In addition, this model can integrate multiple indicators such as blood glucose, blood lipids, and HbA1c to make a comprehensive diagnosis, making up for the defect that traditional methods are difficult to process high-dimensional data.

4. AI Innovation 2: Deep Learning and Convolutional Neural Networks

Deep learning, a subfield of machine learning, uses artificial neural networks to model complex patterns in data. A neural network is a computational architecture inspired by biological neurons, consisting of layers of interconnected nodes that transform input data through weighted connections and nonlinear activation functions. Convolutional neural networks (CNNs) are a specialized type of neural network designed to process grid-like data such as images. CNNs use convolutional filters and pooling operations to automatically extract hierarchical spatial features—from simple edges to complex pathological structures—making them particularly powerful for medical image analysis.

CNNs directly improve both the accuracy and reliability of medical testing. For accuracy, CNNs automatically extract precise pathological features (e.g., lung nodules, tumor contours) from medical images, reducing subjective errors in manual reading (Litjens et al., 2017; Tomassini et al., 2022). For reliability, transfer learning enables stable model training with limited medical data, while outputting confidence intervals to quantify result certainty, addressing the unreliability of traditional threshold methods in complex cases (Goodfellow et al., 2016).

Transfer learning is an important technology to assist CNN training. The model can reuse pre-trained networks and complete training with a small number of samples, which perfectly solves the problem of insufficient medical data in rare disease diagnosis (Goodfellow et al., 2016).

Concrete applications demonstrate the clinical impact of these techniques:

- Lung cancer screening: Ardila et al. (2019) developed a deep learning system trained on 42,290 low-dose CT scans. Their model achieved an area under the curve (AUC) of 0.944 on the National Lung Screening Trial dataset, outperforming six radiologists in sensitivity (94.4% vs. 93.3%) and reducing false positives by 11%.

- Diabetic retinopathy screening: Gulshan et al. (2016) trained a CNN on 128,175 retinal images to detect referable diabetic retinopathy. In validation, the model achieved 90.3% sensitivity and 98.1% specificity, demonstrating AI's potential for large-scale screening in resource-limited settings.

- Transfer learning for rare diseases: Esteva et al. (2017) showed that CNNs pretrained on ImageNet (1.2 million images) could be fine-tuned with as few as 1,000 dermatology images to achieve dermatologist-level classification of skin cancer.

However, deep learning has the risk of shortcut learning. If the training data is biased, the model will learn irrelevant marks instead of real pathological features. Therefore, high-quality and diverse data is the premise of accurate and reliable diagnosis.

5. Extended AI Applications in Medical Detection

AI can be extended to more medical scenarios beyond image and indicator detection. Audio signals such as lung sounds and heart sounds can be converted into spectrograms through short-time Fourier transform, and then classified by CNN-based models. Recent studies have shown the feasibility of this approach for non-invasive respiratory disease screening and telemedicine applications (Perna et al., 2023).

For time-series medical data, sequence models have important value. Early LSTM models can capture long-term and short-term context information, which is suitable for analyzing continuous physiological data such as 24-hour electrocardiograms (ECGs) or continuous glucose monitoring. The Transformer model based on self-attention mechanism improves training efficiency and has been widely adopted in analyzing long-term medical records and predicting disease progression (Goodfellow et al., 2016). These models can assist doctors in interpreting longitudinal patient data and forecasting clinical trajectories, further expanding the application scope of AI in medical treatment.

6. Conclusion

Traditional medical testing methods have obvious limitations in dealing with data uncertainty, individual differences and complex data analysis. Bayesian probabilistic models quantify uncertainty and realize reasonable probability reasoning for individuals. Deep learning and CNNs efficiently process medical images and support personalized diagnosis with the help of transfer learning. Extended technologies such as audio classification and sequence models further enrich the application scenarios of AI. The combination of these AI innovations makes medical testing more accurate, reliable and personalized, which is the inevitable development trend of modern precision medicine. In the future, AI will continue to play an important role in improving medical quality and benefiting more patients.

References

[1] American Diabetes Association. (2023). 2. Classification and diagnosis of diabetes: Standards of care in diabetes—2023. Diabetes Care, 46(Supplement 1), S19–S40. https://doi.org/10.2337/dc23-S002.

[2] Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., … & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine, 25(6), 954–961. https://doi.org/10.1038/s41591-019-0447-x.

[3] Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. New England Journal of Medicine, 299(18), 999–1001. https://doi.org/10.1056/NEJM197811022991808.

[4] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056.

[5] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

[6] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., … & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216.

[7] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

[8] Manrai, A. K., Bhatia, G., Strymish, J., & Kohane, I. S. (2016). Medicine’s uncomfortable relationship with math: Calculating positive predictive value. JAMA Internal Medicine, 176(1), 97–98. https://doi.org/10.1001/jamainternmed.2015.6889.

[9] Murphy, K. P. (2012). Machine learning: A probabilistic perspective. MIT Press. https://doi.org/10.7551/mitpress/9780262018029.001.0001.

[10] Perna, D., Bhaumik, S., & Roy, K. (2023). Deep learning for lung sound analysis: A review. Computers in Biology and Medicine, 157, 106791. https://doi.org/10.1016/j.compbiomed.2023.106791.

[11] Tomassini, S., Falcionelli, N., Sernani, P., Burattini, L., & Dragoni, A. F. (2022). Lung nodule diagnosis and cancer histology classification from computed tomography data by convolutional neural networks: A survey. Computers in Biology and Medicine, 146, 105691. https://doi.org/10.1016/j.compbiomed.2022.105691

[12] Wang, W., Xu, Y., Gao, R., Lu, R., Han, K., Wu, G., & Tan, W. (2020). Detection of SARS-CoV-2 in different types of clinical specimens. JAMA, 323(18), 1843–1844. https://doi.org/10.1001/jama.2020.3786.