Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3 (2025)

Enhancing the Medical Diagnosis System and Treatment by Counterfactual Diagnostic Algorithm

Dr. K. Rama Krishna[1], P Prabakaran[2], Prof. Monali Shetty[3], Vikash Sawan[4], Dr. Hari Jyothula [5], Dr.S.Suma Christal Mary Sundararajan [6]

[1]Department of Information Technology, Vasavi College of Engineering, Hyderabad, India.
k.ramakrishna@staff.vce.ac.in

[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India.
pprabakaran@kluniversity.in

[3]Department of Computer Engineering, Fr. Conceição Rodrigues College of Engineering, Mumbai,India.
shettymonalin@gmail.com

[4]Department of Computer Engineering & Applications, GLA University, Mathura, Uttar Pradesh, India.
vikash.sawan@gla.ac.in

[5] Department of Computer Science and Engineering, Aditya University, Surampalem, India. dr.jyothulahari@gmail.com

[6] Department of Information Technology, Panimalar Engineering College, Chennai, India. sumasheyalin@gail.com

Article History:

Received: *03-08-2024*

Revised: *24-09-2024*

Accepted: *07-10-2024*

Abstract:
Clinical diagnosis and decision-making stand to be radically altered by machine learning. By identifying the underlying medical conditions, doctors may better explain their patients' symptoms via medical diagnosis. There is a substantial correlation between a patient's symptoms and medical history, yet current diagnostic algorithms only detect illnesses that are associated with each other. Suboptimal or even harmful diagnoses may arise from this failure to differentiate correlation from cause, as this article demonstrates. To get around this, they create novel algorithms for counterfactual diagnostics and reframe diagnosis as a work of counterfactual inference. In this work, researchers demonstrate that this method greatly enhances the reliability and security of the subsequent diagnosis while getting closer to the way doctors think about health problems. Using a battery of clinical scenarios, they evaluate the counterfactual method against 45 medical professionals, the gold standard Bayesian diagnosis system, and other algorithms. Compared to the counterfactual approach, which reaches expert clinical accuracy, the Bayesian algorithm obtains correctness, ranking in the top 26% of clinicians in the study unit. Changing this querying strategy alone yields this improvement, eliminating the need for any further model enhancements. Based on these results, the use of machine learning in medical diagnosis is incomplete without counterfactual thinking.

Keywords: Decision Making, Machine Learning, Counterfactual Diagnosis, Health Problems, Clinical Accuracy.

1. Introduction

The fundamental problem for healthcare systems worldwide is to provide an accessible and accurate diagnosis. Misdiagnosis occurs in around 4% of outpatient visits in India annually. Misdiagnoses of major medical illnesses affect a disproportionately large number of people; one-third of these cases

lead to severe patient injury, and around 21% of these individuals get an incorrect diagnosis at the primary care level.

Machine learning and artificial intelligence have recently grown into potent resources for tackling complicated challenges across many fields. Machine learning-aided diagnosis in particular has great potential to transform healthcare by making use of vast amounts of information about patients to provide accurate and tailored diagnoses. In differential diagnosis, when a patient's symptoms might have numerous origins, diagnostic algorithms have failed to match the accuracy of human physicians, despite resurgent business interest and extensive research. Because of this, it's important to know why current methods have such a hard time with differential diagnosis. By analysing the correlation between a patient's symptoms and medical history, all current diagnostic algorithms, including Deep Learning and Bayesian model-based methods, use associative inference to diagnose illnesses. On the other hand, while making a diagnosis, clinicians often look for conditions that would best explain the patient's symptoms. Among the several inference systems that have been identified, associative inference stands out as the most basic. At the very top of this hierarchy is counterfactual inference, which enables one to assign causal explanations to facts. The main point is that making a diagnosis boils down to a game of counterfactual inference. The inability to separate correlation from causation severely limits the precision of associated diagnostic algorithms, which can lead to less-than-ideal or even harmful diagnoses. To overcome this, the authors suggest a causal concept of diagnosis that is more in line with how clinicians make decisions and further get fresh counterfactual diagnostic methods to support this method.

Utilising an experimental set of 1672 clinical vignettes, they evaluate the accuracy level of the proposed counterfactual algorithms against a cohort of 45 clinicians and a cutting-edge associative diagnostic algorithm. The algorithm which is in associative places in the topmost 49% of physicians in the study cohort with an accuracy of 72.53%, compared to the doctors' average diagnosis accuracy of 71.41% in the studies. In contrast, the counterfactual algorithm attains expert clinical accuracy and ranks in the highest 26% of the cohort with an average accuracy rating of 77.27%. Because diagnostic mistakes are more prevalent and can have more severe consequences for uncommon illnesses, these advancements stand out. When comparing the associative method with the counterfactual approach, they find that the latter provides much higher diagnosis accuracy for uncommon and very rare disorders.

The important thing is that both the associative and counterfactual algorithms use the same illness model; the only difference is the querying mechanism. This means that both algorithms improve upon each other. Because learning illness models consumes a lot of resources, backwards compatibility is crucial. Current Bayesian diagnostic models, whether in or out of the medical field, may therefore benefit from the techniques as an instant improvement.

2. Diagnosis

The guiding ideas and presumptions of the present algorithmic diagnostic methodology are outlined below. As a bonus, they detail the scenarios where causal confounding causes this technique to fail and provide a set of guidelines for creating diagnostic procedures that avoid these problems. Lastly,

let's suggest two novel diagnostic algorithms that are founded on the ideas of adequate and necessary causality, using these principles as a basis.

2.1 Associative Diagnosis

The use of a method θ to predict the probability of a fault element D given results ε has been associated with a model-based diagnosis from its official description.

$$\boldsymbol{P(D|\varepsilon;\theta)} \qquad (\mathbf{1})$$

In medical terminology, D stands for a disease or disease, while ε may refer to symptoms, test results, and pertinent medical history. Potential illnesses are rated according to their likelihood when diagnosing over several possibilities, as in a differential diagnosis. There are two main types of model-based diagnostic algorithms: those that discriminate, which model the conditional probability distribution for diseases D given input characteristics E (1), and those that generate, which model the previous distribution of illnesses and results and calculate the posterior using Bayes rule.

$$\boldsymbol{P(D|\varepsilon;\theta) = \frac{P(D|\varepsilon;\theta)P(D;\theta)}{P(\varepsilon;\theta)}} \qquad (\mathbf{2})$$

Generative models are usually Bayesian networks, while discriminative diagnostic models are examples of deep learning and neural networks.

The methods used in this technique are similar to the way physicians make diagnoses. The search for underlying causes of a patient's symptoms has long been considered an essential part of the diagnostic procedure. In this case, for instance, Habitual abduction is the process by which hypotheses are generated. A doctor's ability to diagnose a patient's symptoms is dependent on her familiarity with potential reasons. Similarly, a diagnostic is defined as an examination of the source or origin of an issue, ailment, or problem. In other words, a doctor will try to diagnose the patient's symptoms by looking at the patient's history and current health status to identify the illnesses that are most likely to be causing those symptoms. The following is a proposed causal definition of diagnosis:

- The process of determining, using a patient's medical history, which illnesses are most likely to be causing their symptoms.

There are a lot of literary works that say causal reasoning should be at the heart of diagnosis, yet as far as they are aware, no model-based diagnostic approaches use contemporary causal analysis methods. In most causal contexts, employing the posterior to determine causal linkages might result in false findings due to confounding. Figure 1 (a) displays a condition D that causes symptom S. In this case, addressing D may relieve S. Figure 1 (b) shows that variable R complicates the relationship between S and D; for instance, R may be a hereditary factor that raises a patient's risk of getting illness D and suffering symptom S. In this case, there is a substantial correlation between D and S; nonetheless, a credible diagnosis would not be possible if P and D were to have created symptom S. It is not possible to easily untangle the directed and common sources of symptoms in disorders, as seen in figure 1 (c). Since the posterior cannot distinguish between these many situations, it is inadequate for diagnosing symptoms from patients in all but the most basic conditions, particularly in cases when there are several potential reasons.

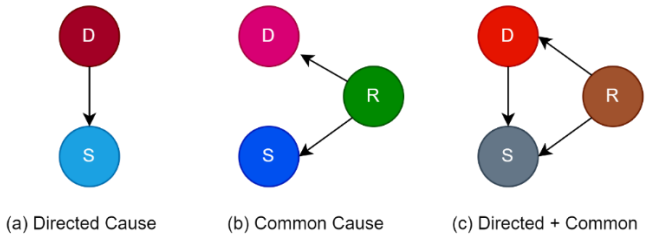

Figure 1 a) Symptom S is caused by disease D; b) D is not the primary source of symptom S, but there is a latent common factor R that correlates them; and c) Both disease D and the presence of the latent similar cause R produce symptom S.

2.2 Guidelines for Making Diagnoses

As an alternate method to associative diagnosis, one might consider the likelihood of causal responsibility or the likelihood that the effect S was caused by the target cause D. This calls for a diagnostic metric M(Dε) to assess the probability that an illness D is the cause of a patient's symptoms based on evidence E. Three very minimum requirements for any such diagnostic tool:

- Consistently, the chance of a disease D eliciting signs should be proportionate to its posterior likelihood (M(Dε) P (D =T ε)).
- No diagnosis can be made for a disease D that does not induce any symptoms, since M(Dε) =0 (causality).
- Simpler diseases that explain more patient symptoms are more probable.

Desiderata is justified as follows. According to Desideratum i), the chance of an illness causing symptoms is proportionate to the possibility of the patient having the condition. If illness D cannot root any of the symptoms from the patients, then D should be ignored (Desideratum ii). Desideratum iii) relies on Occam's razor approach, supporting humble diagnosis with few illnesses to explain symptoms. The posterior only meets the first desiderata, breaching the final two.

2.3 Counterfactual Diagnosis

Utilising counterfactual inference, researchers may measure the probability that a patient's symptoms are caused by a disease. Using a counterfactual, one may determine whether an alternative set of circumstances would have led to a different result. With the given data E = e, they may determine the probability that, in the absence of some hypothetical intervention, they would have seen a different result-$\boldsymbol{\varepsilon = e'}$, which runs contrary to the fact that ε = e. In Pearl's calculus of interventions, the intervention which sets the parameter rX to the value X = x is denoted by do (X = x) Hence the counterfactual probability is expressed as $\boldsymbol{P(\varepsilon = e' | \varepsilon = e}$, do (X = x)). Counterfactuals quantify the effectiveness of an ailment the hypothesis D = T in explaining symptoms evidence S = T by calculating the probability that the symptom is contemporary if the disease were cured (P (S = F S = T, do (D = F)) (do (D = F)). If the likelihood is higher, D = T provides a valid fundamental explanation for the symptom. This probability cannot be expressed as a normal posterior since it pertains to two contradicting S states. Appendix C, explain the calculation of counterfactual probabilities.

The circumstances of this analysis served as inspiration for the two proposed counterfactual diagnostic metrics: anticipated sufficiency and expected disablement. Both measures meet all three requirements from section 2.2 as shown in the first theorem at the conclusion of this section.

First Definition (Predictable Impairment): The predicted impairment of illness D is the range of symptoms that would change if the disease were cured.

$$\mathrm{E}_{dis}(\boldsymbol{D}, \boldsymbol{\varepsilon}) = \sum_{s'} |\boldsymbol{S}_+ \setminus \boldsymbol{S}'_+| \boldsymbol{P}\left(\boldsymbol{s}' \middle| \boldsymbol{\varepsilon}, \mathbf{do}\ (\mathbf{D} = \mathbf{F})\right) \qquad (\mathbf{3})$$

In this case, E represents factual proof and S+ represents positively proven symptoms. Summarise all counterfactual symptom evidence statements. S and S+ represent counterfactual symptoms that are affirmatively demonstrated. do (D = F) represents $\mathbf{D} \rightarrow \mathbf{F}$ counterfactual intervention. The cardinality of the set of symptoms that are factually present but not in the counterfactual symptoms data is denoted by $|\mathbf{S}_+ \setminus \mathbf{S}'_+|$.

The concept of required cause is where the predicted disablement is derived from. If S = T, then D is a required source of S and vice versa. Consequently, the anticipated disability measures both the adequacy of illness D in explaining the symptoms from patients and the probability that treatment D alone would ameliorate those symptoms.

Second Definition (Adequate): D Disease sufficiency is the number of positive symptoms predicted to exist after excluding all other plausible causes.

$$\mathrm{E}_{suff}(\boldsymbol{D}, \boldsymbol{\varepsilon}) \coloneqq \sum_{s'} \boldsymbol{S}'_+ | \boldsymbol{P}(\boldsymbol{s}' | \boldsymbol{\varepsilon}, \mathbf{do}\ (\mathbf{Pa}\ (\boldsymbol{S}_+) \setminus \boldsymbol{D} = \boldsymbol{F})) \qquad (\mathbf{4})$$

Where the aggregate of all potential confounding symptoms indicates in the hypothetical symptom condition, the symptoms that are positively demonstrated are denoted by $\boldsymbol{s'}\ \boldsymbol{and}\ \boldsymbol{s'}$. In this context, Pa(S+) D refers to the set of all directly causative variables for the set of affirmatively proved symptoms that do not include illness D, and do $\boldsymbol{Pa}(\boldsymbol{S}_+) \setminus \boldsymbol{D}$stands for the counterfactual interference that sets all $\boldsymbol{Pa}(\boldsymbol{S}'_+ \setminus \boldsymbol{D}) \rightarrow \boldsymbol{F}$. $\boldsymbol{\varepsilon}$stands for the collection of all evidence based on facts. In the case of hypothetical symptoms, $|\boldsymbol{S}'_+|$ indicates the cardinality of the collection of all possible symptoms.

The expectation of sufficiency is based on the concept of adequate cause, which states that if the existence of D suggests the later occurrence of S, then D is an adequate cause of S. Nevertheless, the anticipated sufficiency is not met since S could exist for various causes; that is, S's presence does not automatically indicate that D occurred before. In most cases, symptoms may be adequately explained by illnesses. Following the execution of counterfactual actions, every possible reason for the symptoms has been eliminated, leaving just D as a potential cause. They next isolate its impact as an adequate reason in the theoretical framework.

Theorem 1: qualities of predicted disability and expected sufficiency as diagnostic criteria. The three requirements from section 2.2 are met by anticipated incapacity and anticipated sufficiency.

3. Methodology

In this part, the developers provide the statistical illness models that will be used to evaluate the diagnostic metrics discussed before. Simpler formulas for the predicted sufficiency and disablement in these models are subsequently derived.

3.1 Diagnostic Structural Causal Models

Bayesian Nets represent the links between hundreds of risk factors, illnesses and symptoms; researchers apply these models to the study. BNs are often used in diagnostics because they can be understood and specifically represent the relationships between variables that are necessary for causal and hypothetical analysis. In these models, illnesses, symptoms, and risk factors are usually shown as true or false binary nodes. The conventional integer notation is used to represent true as 1 and false as 0, respectively.

Directional acyclic graphs (DAGs) and joint probability distributions across all nodes that factorise about the DAG architecture define BNs. Assuming a fixed arrow exists between two nodes X and Y, say that X is a parent of Y and that Y is a child of X. Assuming a directed link exists between two nodes Y and Z, say that Z is an ancestor of Y. Figure 2 (a) shows a basic example BN that models' illnesses, indicators, and risk factors.

The INTERNIST-1, QMR, and PATHFINDER systems are among the first examples of BN disease models; these systems associated with noisy-OR networks had just illness and symptom nodes, or BN2Onetworks. Figure 2 (a) shows that three-layer BNs have recently supplanted these two-layer models. In addition to including illness risk indicators, these models reduce reliance assumptions. The models for which the results will be derived are easily generalisable to others, regardless of the complexity of their relationships.

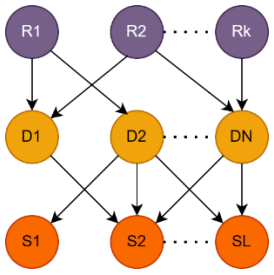

Figure 2 (a) Three-layer BN Method

In causal inference, Structural Causal Models (SCMs), sometimes called Functional Causal Models or Structural Equation Models, supersede BNs. A lot of research and practical experience with SCMs has gone into understanding their relationship to other methods like BNs and probabilistic graphical models.

An essential feature of SCM is that they depict all variables as probabilistic combinations of their immediate causes and an unobserved external noise component that stands in for all non-model causes. There is a probability distribution across the observed variables due to the fact that the state of the noise component is unknown.

3.2 Dual-OR Noisy Diagnostic Networks

Building disease models sometimes involves making extra assumptions about the model beyond what is suggested by the DAG architecture. Among these, noisy-OR models are the most popular. Because they represent common sense assumptions about the relationships between symptoms and illnesses, noisy-OR models find widespread use in medical modelling. Furthermore, they provide effective reasoning and learning, and they enable a multitude of characteristics that scale linearly with the network size to characterise massive BNs.

A parent Di may activate their kid S using the noisy-OR assumption if two conditions are met: (i) the parent is turned on, (ii) $\boldsymbol{D_i}$ = 1, and (iii) the activation procedure does not fail at random. There is no correlation between any of the other model parameters and the failure probability, which is sometimes written as $\boldsymbol{\lambda D_i, S}$.

The noisy-OR component activates the kid if either parent successfully activates it. The value of S is the Boolean OR operation of its parents' functions of activation, $\boldsymbol{s = v_i f(d_i, u_i)}$, where $\boldsymbol{f(d_i, u_i) = d_i \wedge \overline{u_i}}$, $\wedge$, denotes the Boolean AND function, $\boldsymbol{d_i \epsilon \{0, 1\}}$ is the state of a parent, and $\boldsymbol{u_i \epsilon \{0, 1\}}$ is a latent noise variable with a chance of failure $\boldsymbol{P(u_i = 1) = \lambda D_i, S}$. If a symptom can be activated with a single activation and activating a disease does not activate a symptom, then the noisy-OR model intuitively reflects this circumstance.

In order to calculate the predicted sufficiency and disablement for these models, they use the twin-networks approach for calculating counterfactuals. For the purpose of computing counterfactual probabilities using traditional inference methods, this approach integrates real and counterfactual parameters into a single SCM twin network. Abduction remains unsolvable for large SCMs, but this approach drastically reduces the inferential cost of computing counterfactuals.

4. Results

The studies comparing posterior inference with the predicted disablement and sufficiency are detailed here, using the models described above. Presenting the test set, which consists of a group of physicians and a collection of clinical scenarios. These algorithms are then tested on a variety of diagnostic tasks for evaluation.

4.1 Databases and Diagnostic Method

The usage of EHRs is one method for verifying diagnostic algorithms. When diagnostic mistakes lead to incorrectly labelled data, a major drawback of this method is how hard it is to define the ground truth diagnosis. This problem is especially noticeable in differential diagnoses because of the high rates of diagnostic ambiguity and miscommunication, the wide variety of applicant diseases and their associated diagnosis labels, the lack of or incorrect documentation of case details, and biases like the diagnosing doctor's expertise and education. To address these concerns, it is common practice to examine clinical vignettes or simulated diagnostic situations when evaluating doctors.

A clinical vignette comprises a non-exhaustive compilation of evidence, including indicators, medical records, and essential demographic information such as age and birth gender, illustrating a typical patient's manifestation of a problem. This approach has been effective in evaluating physicians as individuals and juxtaposing their precision with symptoms checker technologies; it is also more

resistant to inaccuracies and biases than actual data sets such as electronic health records (EHRs). Practising a disease depending on its established characteristics is more straightforward than doing a differential diagnosis.

A different group of physicians, all with medical training at least up to the general practitioner level, created 1672 clinical vignettes for us to utilise as a test. Every effort has been made to ensure that the symptoms and risk variables align with the statistical illness model. However, to keep the study unbiased, the scenes include any extra clinical details that the clinicians in the studies may have as case notes. Each case narrative is written by one doctor and then double-checked by other physicians to make sure it depicts a real-life diagnostic scenario. The algorithm conceals the actual condition in each vignette and then uses the information from the vignettes to rank all of the modelled diseases and provide a diagnosis. The associative method uses the posterior to rank diseases, whereas the counterfactual algorithms use the predicted disablement or expected sufficiency. A partly sorted list of potential illnesses is the independent differential diagnosis that doctors provide.

To guarantee that any modification in diagnostic correctness is related to the rank query used, the associative and counterfactual algorithms use the same disease models in all investigations. The BN is defined by a coalition of medical professionals and epidemiologists. Epidemiological data are used to determine the prior likelihood of diseases and risk variables, while several independent medical specialists are interviewed to determine the conditional probabilities.

4.2 Comparison of Associative and Counterfactual Rankings

In the first experiment, we evaluate the diagnostic accuracy of rating illnesses by posterior anticipated disablement and expected sufficiency. The top-k correctness is defined as the ratio of the 1672 diagnosis vignettes in which the actual disease is included in the top-k rankings.

. For each of the 1672 vignettes, the top-k ranked diagnoses are computed, with k = 1..., 20. The findings are shown in Figure 3. The accuracy for all k on the test set is about the same while using the anticipated disablement and expected sufficiency, hence they only provide the consequences for the expected adequacy for clarity's sake.

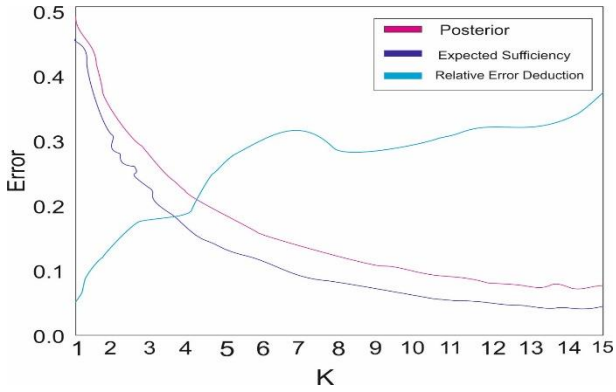

Figure 3 Top k Ranked Accuracy Findings.

The counterfactual method outperforms the associative algorithm by 2.6% when k = 1, giving the top-ranking illness. Both methods function differently for k > 1, with the counterfactual strategy significantly outperforming the associative algorithm in terms of error rate reduction. With k > 5, the

counterfactual approach outperforms the associative algorithm in reducing misdiagnoses by around 31%. This indicates that although the posterior does a decent job of finding the best candidate illness, the counterfactual ranking does a much better job of identifying the next most probable diseases. When making a differential diagnosis, triage, and deciding on the best course of therapy, these secondary candidate disorders take on added significance.

Finding the genuine disease's place in the rankings is a quick and easy way to compare two rankings. Among all 1672 vignettes, we discovered that in 24.8% of cases, the real sickness was rated higher by the counterfactual algorithm compared to the associative method, and in 2% of cases, it was ranked lower. Compared to the associative method's 3.82 plus or minus 5.26, the genuine illness is rated at 3.17 plus or minus 4.5 using the counterfactual approach, which is a significant improvement.

The vignettes are organised according to the genuine disease's past incidence rates: extremely frequent, common, uncommon, rare, and very rare. For both common and uncommon illnesses, the counterfactual algorithm outperforms the associative method. However, for very uncommon diseases, the enhancement is much more pronounced, with the counterfactual approach attaining a superior ranking for 33 per cent of these vignettes and rare disorders for 29.3 per cent. Since uncommon diseases are notoriously difficult to detect, this development is crucial since they include a wide range of dangerous illnesses for which diagnostic mistakes may have devastating effects.

4.3 Compared to Doctors

In the following study, they evaluate the associative and counterfactual algorithms with a group of 45 medical professionals. Every doctor gets a set of 51 vignettes to work with. They then have to independently diagnose each vignette using a partly ranked list containing k illnesses, wherein the dimensions of the list k are determined by the doctor for each instance. Given a doctor and a set of vignettes that need a diagnosis, the associated and counterfactual algorithms are fed the same data and each one gives back a top-k evaluation, where k is the doctor's size. To evaluate the algorithms' and physicians' accuracy, they match their precision for each vignette. This way, researchers don't have to force clinicians to diagnose a specific number of illnesses. This is significant because, as the diagnostic vignette displays the physicians' level of doubt, the magnitude k of their diagnosis will naturally fluctuate.

The full outcomes for all 45 doctors as well as the predicted sufficiency, posterior, and expected disablement grading processes. The associated and counterfactual algorithms are compared to each doctor's accuracy in Figure 4. For each of the 45 doctors, one can see their average accuracy; this is based on the percentage of vignettes in which they correctly identified the underlying illness. The equivalent algorithm's accuracy in diagnosing the identical vignette and returning differences of the same magnitude as that doctor is displayed against this measure. If both the doctors and the algorithms are quite accurate, then the case sets with the simplest vignettes will have the best results. On the other side, when it comes to more difficult vignettes that both the clinicians and the computer struggle with, the algorithm usually ends up with better accuracy. This indicates that diagnostic algorithms should be considered a supplement to human physicians, with algorithms demonstrating superior performance in scenarios where human error is more prevalent.

The associative procedure achieves a mean accuracy over all trials of 72.53 + or - 2.98%, which is comparable to the average doctor's 71.41 plus or minus 3.02%. The algorithm outperforms 22 out of the doctors, ties with 1 doctor, and underperforms 22 out of the doctors. In comparison to both the typical doctor and the associated algorithm, the counterfactual algorithm obtains an impressive average accuracy of 77.26 plus or minus 2.80%. This puts it in the top 26% of physicians in the cohort. Contrary to fact method outperforms 33 out of the physicians, has a draw percentage of 1, and is less accurate than 11.

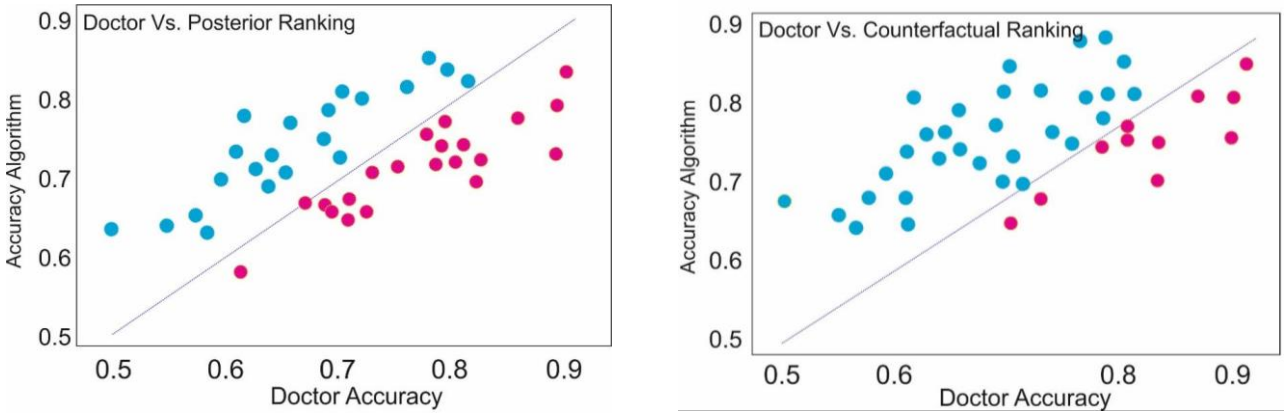

Figure 4 Associated and Counterfactual Algorithms Are Compared to Each Doctor's Accuracy

All things considered; the diagnostic accuracy is much better with the counterfactual approach compared to the associative algorithm. For uncommon illnesses, we find the improvement to be quite striking. In contrast to the counterfactual algorithm's top-quartile performance, the associative algorithm's results are on par with the typical doctors.

5. Discussion with Conclusion

Healthcare systems throughout the world face a formidable obstacle in the form of inadequate access to basic care and the problem of incorrect differential diagnoses. To use machine learning to the advantage and conquer these obstacles, they must first comprehend the diagnostic process and define precisely what they want the algorithms to produce. Diagnostics and associative inference have been confused in previous methods. The former seeks to identify the root of a patient's symptoms, while the latter learns associations between patient records and illness incidences, to ascertain the most probable diseases in the patient's community. In straightforward causal situations with a single illness, this method may work, but it severely limits the accuracy of these procedures. When a doctor must pick between competing disease theories. To get over these limitations, researchers need to reevaluate diagnostic criteria and algorithm design from the ground up.

In this paper, researchers define diagnosis causally differently and contend that diagnosis is essentially an exercise in counterfactual inference. Anticipated disablement and anticipated sufficiency are two counterfactual diagnostic measures that have been devised. To calculate these measures, they have introduced a new class of diagnostic models called twin diagnostic networks. Utilising pre-existing diagnostic models, researchers proved that, in comparison to conventional associative rankings, using these counterfactual metrics to rank illness hypotheses substantially enhances diagnosis accuracy. When it came to expert clinical accuracy, the counterfactual procedure was in the top 26 per cent of

the sample, whereas the associative algorithm was around average. In 29.3% of instances, the counterfactual procedure ranked the genuine illness higher than the associated process; in 33% of these situations, the improvement is especially noticeable for very uncommon diseases, where diagnostic mistakes are usually more numerous and more severe. The best part is that the disease model doesn't even need to be changed to take advantage of this enhancement. Existing Bayesian diagnostic methods, whether in or out of the medical field, may benefit from the algorithm's instant improvement thanks to its backward compatibility.

In contrast to previous efforts that have centred on enhancing model architectures or tapping into new data sources, findings provide a novel way to improve clinical decision systems at the expert level by altering model queries to take causal information into account. These results provide credence to the claim that, in certain fields, human specialists will always be able to outperform machine learning approaches that do not use causal reasoning. The current tests have concentrated on comparing the algorithms to doctors, but future trials might find out how well algorithms work as medical support systems that help doctors by giving them a second opinion when diagnosing patients. The combined diagnosis of a doctor and the method will probably be more reliable than either one alone, as the system seems to complement human physicians by doing better in scenarios that humans have trouble diagnosing.

Reference:

[1] Richens, Jonathan G., Ciarán M. Lee, and Saurabh Johri. "Counterfactual diagnosis." *arXiv preprint arXiv:1910.06772* (2019).

[2] Richens, Jonathan G., Ciarán M. Lee, and Saurabh Johri. "Improving the accuracy of medical diagnosis with causal machine learning." Nature Communications 11.1 (2020): 3923.

[3] Prosperi, Mattia, et al. "Causal inference and counterfactual prediction in machine learning for actionable healthcare." Nature Machine Intelligence 2.7 (2020): 369-375.

[4] Verma, Sahil, et al. "Counterfactual explanations and algorithmic recourses for machine learning: A review." ACM Computing Surveys (2020).

[5] Singla, Sumedha, et al. "Explaining the black box smoothly—a counterfactual approach." Medical image analysis 84 (2023): 102721.

[6] Lin, Junfan, et al. "Towards causality-aware inferring: a sequential discriminative approach for medical diagnosis." IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).

[7] Wu, Hang, et al. "Clinical decision making under uncertainty: a bootstrapped counterfactual inference approach." BMC Medical Informatics and Decision Making 24.1 (2024): 1-15.

[8] Thiagarajan, Jayaraman J., et al. "Training calibration-based counterfactual explainers for deep learning models in medical image analysis." Scientific Reports 12.1 (2022): 597.

[9] AlJalaud, Ebtisam, and Manar Hosny. "Counterfactual Explanation of AI Models using an Adaptive Genetic Algorithm with Embedded Feature Weights." IEEE Access (2024).

[10] Temraz, Mohammed, and Mark T. Keane. "Solving the class imbalance problem using a counterfactual method for data augmentation." Machine Learning with Applications 9 (2022): 100375.

[11] Verma, Sahil, John Dickerson, and Keegan Hines. "Counterfactual explanations for machine learning: A review." arXiv preprint arXiv:2010.10596 2 (2020): 1.

[12] Kelly, Luke, et al. "Evolving Visual Counterfactual Medical Imagery Explanations with Cooperative Co-evolution using Dynamic Decomposition." Proceedings of the Genetic and Evolutionary Computation Conference Companion. 2024.

[13] Atad, Matan, et al. "Counterfactual Explanations for Medical Image Classification and Regression using Diffusion Autoencoder." arXiv preprint arXiv:2408.01571 (2024).

[14] Dai, Xinyue, et al. "Counterfactual explanations for prediction and diagnosis in XAI." Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. 2022.

[15] Cheng, Furui, Yao Ming, and Huamin Qu. "Dece: Decision explorer with counterfactual explanations for machine learning models." IEEE Transactions on Visualization and Computer Graphics 27.2 (2020): 1438-1447.

[16] Huang, Hao, Emetis Niazmand, and Maria-Esther Vidal. "Hybrid AI Approach for Counterfactual Prediction over Knowledge Graphs for Personal Healthcare." Artificial Intelligence and Data Science for Healthcare: Bridging Data-Centric AI and People-Centric Healthcare. 2024.

[17] Chou, Yu-Liang, et al. "Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications." Information Fusion 81 (2022): 59-83.

[18] Mertes, Silvan, et al. "Ganterfactual—counterfactual explanations for medical non-experts using generative adversarial learning." Frontiers in artificial intelligence 5 (2022): 825565.

[19] Singh, Chaitanya, et al. "Applied machine tool data condition to predictive smart maintenance by using artificial intelligence." International Conference on Emerging Technologies in Computer Engineering. Cham: Springer International Publishing, 2022..

[20] Dash, Saloni, Vineeth N. Balasubramanian, and Amit Sharma. "Evaluating and mitigating bias in image classifiers: A causal perspective using counterfactuals." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.

[21] Son, Jaemin, et al. "An interpretable and interactive deep learning algorithm for a clinically applicable retinal fundus diagnosis system by modelling finding-disease relationship." Scientific Reports 13.1 (2023): 5934.

[22] Barraza, Joaquín Figueroa, Enrique López Droguett, and Marcelo Ramos Martins. "FS-SCF network: Neural network interpretability based on counterfactual generation and feature selection for fault diagnosis." Expert Systems with Applications 237 (2024): 121670.

[23] Zhang, Baoliang, et al. "Counterfactual inference graph network for disease prediction." Knowledge-Based Systems 255 (2022): 109722.

[24] Sun, Zhaohong, et al. "Adversarial reinforcement learning for dynamic treatment regimes." Journal of Biomedical Informatics 137 (2023): 104244.

[25] Patro, R. Azhagumurugan, R. Sathya, K. Kumar, T. R. Kumar and M. V. S. Babu, "A hybrid approach estimates the real-time health state of a bearing by accelerated degradation tests, Machine learning," 2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2021, pp. 1-9, doi: 10.1109/ICSTCEE54422.2021.9708591