

Using Fuzzy Cognitive Maps to Clarify Causality vs Correlation in Biostatistics and Bioinformatics

Peter P Groumpos*

Emeritus Professor, University of Patras, Greece

ISSN: 2578-0247



***Corresponding author:** Peter P Groumpos, Emeritus Professor, University of Patras, Greece

Submitted:  August 28, 2026

Published:  October 23, 2026

Volume 4 - Issue 2

How to cite this article: Sharon L Burton*. The Role of Cybersecurity in Bioinformatics for Aviation and Aerospace Safety. Open Acc Biostat Bioinform. 4(2). OABB.000583. 2026.
DOI: [10.31031/OABB.2026.04.000583](https://doi.org/10.31031/OABB.2026.04.000583)

Copyright©: Peter P Groumpos. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract

Modern biomedical research relies heavily on data-driven decision-making, elevating the distinction between statistical correlation and true causality to a foundational challenge in biostatistics and bioinformatics. In high-dimensional datasets containing millions of variables—such as genetic variants, transcriptomic profiles, clinical phenotypes, and environmental exposures—identifying meaningful relationships is highly complex. Emerging researchers frequently conflate predictive correlations with causal mechanisms. Misinterpreting these statistical associations can lead to erroneous biological conclusions, ineffective therapeutic interventions, and flawed clinical recommendations. To address this, this paper clarifies the conceptual and mathematical boundaries between correlation and causality within biomedical frameworks, evaluating the utility of Fuzzy Cognitive Maps (FCMs) as an interpretable methodology for structuring and communicating causal hypotheses. As semi-quantitative networks, FCMs model biological components as nodes and their directed interactions as weighted edges, successfully capturing the direction, sign, and approximate magnitude of hypothesized influences. This approach effectively accommodates feedback loops and integrates expert knowledge with empirical data, bridging the gap between purely statistical associations and mechanistic biological reasoning. However, while FCMs serve as powerful tools for hypothesis generation and system visualization, they do not replace formal structural causal inference methodologies. Ultimately, an FCM functions as a formal representation of a hypothesized causal network rather than an autonomous mathematical mechanism to prove causality from observational data alone.

Keywords: Fuzzy cognitive maps; Causality; Correlation; Biostatistics; Bioinformatics; Systems biology; Causal modeling; Biological networks

Introduction

Biostatistics and bioinformatics increasingly rely on large and heterogeneous datasets [1,2]. Advances in high-throughput sequencing, transcriptomics, proteomics, metabolomics, electronic health records, medical imaging, and wearable technologies have created new opportunities for studying biological systems. However, the growing complexity and dimensionality of biomedical data create a fundamental challenge: identifying relationships between variables is much easier than determining whether those relationships are causal [3,4]. For example, a bioinformatics analysis may identify a strong association between gene expression and disease progression. Although this relationship may be biologically meaningful, several explanations are possible. The gene may directly influence disease progression, disease progression may alter gene expression, a third factor may affect both, or the association may result from confounding, selection bias, measurement error, population structure, or other sources of bias.

This challenge is commonly summarized by the principle that correlation does not imply causation [5-9]. While simple in theory, distinguishing correlation from causality in high-dimensional biomedical systems is difficult. Modern causal inference provides formal frameworks for addressing this problem, including structural causal models [10], directed acyclic graphs, potential-outcomes frameworks, interventions, and counterfactual reasoning.

Pearl, [10,11] for example, emphasized that causal analysis requires assumptions about the underlying data-generating process rather than relying only on observed statistical dependencies.

Fuzzy Cognitive Maps (FCMs) provide a complementary perspective. Originally introduced by Kosko, FCMs can represent complex systems where concepts such as inflammation, gene expression, oxidative stress, drug exposure, and disease severity are connected through positive or negative relationships of varying strengths. (More historical information on FCMs is provided in Section 4.) The value of FCMs is not that they automatically transform correlations into causal relationships, but that they make causal assumptions explicit. By creating an FCM, researchers must specify which concepts are believed to influence others, in what direction, and with approximately what strength. This process can reveal assumptions that might otherwise remain hidden within statistical models or computational pipelines.

The objective of this paper is to examine how FCMs can help clarify the distinction between correlation and causality in biostatistics and bioinformatics. It develops a conceptual framework for integrating FCMs with statistical association analysis and formal causal inference while emphasizing the limitations of FCM-based causal reasoning. The usefulness of FCMs in this context is not that they automatically transform correlations into causal relationships. Rather, their value lies in making causal assumptions explicit. A researcher who creates an FCM must state which concepts are thought to influence which others, in what direction, and with approximately what strength. This explicit representation can expose assumptions that might otherwise remain hidden inside statistical models or computational pipelines. After the introduction the sections of the paper are outlined as follows. In section 2 the important scientific terms of Correlation and Causality in modeling complex dynamical systems are defined and their difference is outlined. Section 3 explains why there is a difference between Biostatistics and Causality. In addition, an explanation why Biostatistics and Bioinformation are used extensively in health sciences. In section 4 the new scientific discipline of Fuzzy Cognition Maps (FCM) is introduced and briefly mathematically explained. Then section 5 discusses why FCMs are relevant to Biostatistics and Bioinformatics. In section 6 the way to distinguish between Correlation and Causality is scientifically discussed. In section 7 another interesting question is raised and answered. In seven (7) simple but well-formulated steps we can integrate FCMs with Biostatistics Data and arrive at very useful results. In section 8 a major scientific issue is addressed, the distinction between an FCM representation of causality and a formal causal model.

While section 9 provides some Advantages of FCMs in Biomedical Research and thus in Biostatistics and Bioinformatics. Section 10 provides a short summary of the paper and then introduces and discusses briefly some interesting and future research directions. In the end of this section (10) some interesting thoughts are raised challenging today's. Finally, conclusions are provided in section 11. The paper provides a good number of references which are a good start for the needed humanized and wise vision of the whole world.

Correlation and Causality in Dynamical Complex Systems Research

Correlation and causality are two of the most critical concepts in the study of dynamic complex systems, such as ecosystems, health sciences, the climate, the human brain, and financial markets [5-9]. In these systems, classical statistical correlation often fails to reveal true causal relationships due to non-linear interactions, feedback loops, and time lags. This limitation arises because correlation is primarily suited for linear systems, whereas most dynamical systems are inherently non-linear. Nonetheless, both concepts are extensively applied in biological and healthcare systems, which form the central focus of this paper.

Correlation

Correlation describes a statistical relationship between variables [5]. When two variables change together, a correlation coefficient can quantify the strength and direction of their association. For example, people who exercise more often may have lower blood pressure, demonstrating a correlation between exercise and blood pressure. However, correlation alone does not prove that exercise causes lower blood pressure; other factors could be involved, such as age, diet, smoking, or general health. Correlation is valuable because it provides a starting point for scientific investigation [12]. In high-dimensional bioinformatics or biostatistics, correlation-based analyses can reveal gene clusters, identify candidate biomarkers, detect co-expression patterns, and generate hypotheses about biological pathways [13].

However, correlation alone does not specify the mechanism producing the relationship.

Suppose variables X and Y are correlated. At least several possibilities exist:

- i. X is related to Y.
- ii. Y is related to X.
- iii. A third variable Z is related to both X and Y.
- iv. X and Y influence each other through feedback.
- v. The relationship is partly or entirely due to random variation.

Consequently, a high correlation coefficient should be interpreted as evidence of statistical dependence, not as definitive evidence of a causal relationship [7]. We should keep in mind that correlation does not imply causality.

Causality

Causal reasoning asks a fundamentally different question: What would happen to an outcome if an intervention changed to a particular variable? [7-9,14]. For example, instead of asking whether gene expression X is associated with disease severity Y, a causal question might ask: What change in disease severity would result if gene expression X were experimentally altered while relevant confounding factors were controlled? This distinction is central to causal inference. Formal causal frameworks distinguish

observational associations from the effects of interventions and counterfactual outcomes. For example, Pearl's structural causal model framework [10] provides a mathematical language for expressing assumptions about causal relationships and answering causal and counterfactual questions (Proceedings of Machine Learning [11]).

In biomedical research, establishing causality may require randomized experiments, longitudinal designs, natural experiments, instrumental variables, Mendelian randomization, causal graphical models, or other identification strategies, such as Artificial Intelligence (AI) and Fuzzy Cognitive Maps (FCMs). However, any AI- or FCM-based approach must operate alongside—rather than instead of—these established methods. It is crucial for scientists to accurately distinguish between correlation and causality [7,8]. To achieve this, we must recognize that determining causal relationships is rarely straightforward.

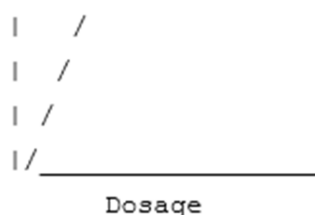
Is there any difference between biostatistics and bioinformatics? Why are both of them used in health science?

This is a complex question to answer. Biostatistics and

bioinformatics are vital to life sciences, playing crucial roles in data analysis, interpretation, and decision-making. Although both fields analyze vast amounts of biological data, they differ in their methodologies and applications. Biostatistics focuses on applying traditional probability and statistical methods to study design, data analysis, and biological inference. In contrast, bioinformatics leverages computational tools and algorithms to analyze and interpret biological data, focusing on molecular-level processes and systems biology. Crucially, researchers in this field must distinguish between correlation and causality. The following sections explore several critical scientific issues inherent to both disciplines.

The illusion of linearity in complex biological systems: A common pitfall for early-career researchers is relying on statistical tools that assume linear relationships. Standard correlation coefficients, such as Pearson's ρ , only measure the strength of a straight-line association. However, human pathophysiology is rarely linear; instead, it is governed by thresholds, feedback loops, and U-shaped curves (Figure 1).

Linear Assumption (Pearson)



Non-Linear Reality
(Biological Truth)

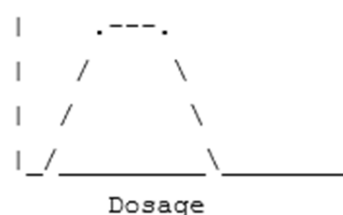


Figure 1: Linear assumption for correlation vs the nonlinear one for biological systems.

The failure of linear correlation: Consider a novel therapeutic compound. At a low dose, it has no effect; at a medium dose, it successfully reduces tumor volume. However, at a high dose, it becomes highly toxic, causing cellular stress that accelerates tumor growth. If a researcher applies a standard linear correlation test to this U-shaped dataset, the positive and negative slopes will mathematically cancel each other out. As a result, the software will report a correlation coefficient near zero (ρ approx. 0). An inexperienced researcher might then mistakenly abandon a highly effective drug simply because their analytical tool assumed the biological system operated in a straight line.

Why causal inference matters in biostatistics?: Because many major biostatistical questions are inherently intervention-focused:

- Does this drug reduce mortality?
- Does this treatment improve survival?

- Does a screening program reduce deaths?
- What would happen to this patient if we gave them treatment A instead of treatment B?

These are all causal questions. For example, imagine a scenario where: Treatment $\rightarrow$ Lower Mortality. If we simply observe that treated patients have lower mortality rates, we still do not know whether the treatment actually caused the reduction. It is entirely possible that younger or healthier patients are simply more likely to receive the treatment in the first place. This distortion is known as confounding. Biostatistics has developed an extensive toolbox to distinguish between a mere association and a genuine treatment effect. This includes randomized controlled trials, statistical adjustment, stratification, propensity scores, instrumental variables, longitudinal models, survival analysis, and causal diagrams, alongside other clinical data analytics. Importantly, statistics is deeply intertwined with causal inference. While causal reasoning can exist without statistics—such as through mechanistic

reasoning or a single, carefully controlled experiment, estimating causal effects from data almost always requires statistical inference.

Why does bioinformatics seem more “correlation-oriented”?: Bioinformatics often focuses on correlation because it frequently begins with high-dimensional observational data [15].

Examples of High-Dimensional Data, 20,000 gene-expression measurements.

- a) Millions of genetic variants.
 - b) Thousands of proteins.
 - c) Single-cell expression profiles.
 - d) Metabolomics measurements.
- a) **You might ask:** Which genes are associated with cancer? Or which genes have similar expression patterns? Or which genomic features predict drug response? Correlation/association is extremely useful for these exploratory questions.
- b) **Another example:** if Gene A expression $\uparrow \leftrightarrow$ Gene B expression $\uparrow$ that may suggest that the genes participate in the same pathway or are co-regulated.
- c) **But again:** Gene A correlates with Gene B, doesn't establish: Gene A causes Gene B.

While bioinformatics utilizes correlation for initial discovery, traditional biological experimentation remains essential to elucidate underlying mechanisms and prove causality.

Fuzzy Cognitive Maps

Let us see how the new theoretical discipline of Fuzzy Cognitive Maps (FCM) can be of some help in addressing several difficult but challenging issues of Biostatistics and Bioinformatics. Especially when are related to both [16-19].

History Of FCM

A Fuzzy Cognitive Map (FCM) is a mathematical modeling tool that combines fuzzy logic and neural networks to represent and analyze complex causal systems [20,21]. In other words, it is a computational methodology capable of examining situations where the human thinking process involves fuzzy or uncertain descriptions. An FCM offers a graphical representation through a signed directed graph with feedback, consisting of nodes and weighted arcs. The nodes of the graph represent concepts used to describe the relations and behavior of a system in a simple, symbolic way through cause and effect. These nodes are connected by signed and weighted arcs, which represent the causal relationships existing between the concepts (Figure 2). Each concept C_i (variable) is characterized by a value calculated either by transforming a fuzzy value or by fitting a numeric value into the desired interval, $[0,1]$. The values of the interconnection weights are initially defined linguistically by experts and then transformed into values within the interval $[-1,1]$ using a specially designed algorithm [20]. In this way, FCMs embody the accumulated knowledge and experience of experts who understand how the system behaves under different circumstances (Figure 2).

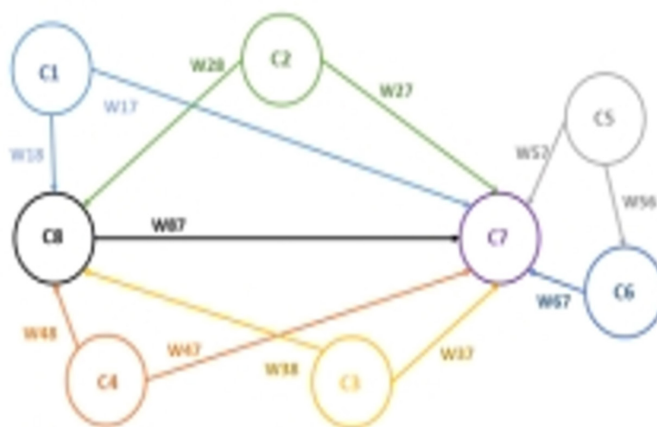


Figure 2: Graphical representation of fuzzy cognitive maps.

FCMs were first introduced by Kosko in 1986 to represent causal relationships between concepts and analyze inference patterns [21,22]. Ten years earlier, in 1976, Robert Axelrod had introduced the concept of Cognitive Maps [23] as a formal framework for representing social scientific knowledge and modeling decision-making in social and political systems. Kosko later integrated computational elements into this framework to develop FCMs. In 1996, a research team led by Prof. Groumpos at the Laboratory of Automation and Robotics (LAR) at the University of Patras advanced the theory further, expanding FCMs through a

significant number of new theoretical accomplishments [20,24-26]. FCMs utilize the knowledge and experience of experts, offering an alternative problem-solving methodology that mirrors human cognition. This is accomplished via a conceptual procedure capable of incorporating ambiguous, fuzzy descriptions [5-9,16-21,23,24,27,28].

Basics of FCM

A Fuzzy Cognitive Map (FCM) is a directed graph in which nodes represent concepts and edges represent influences between

concepts (Figure 2). Each edge is assigned a weight representing the approximate strength and direction of the relationship.

An FCM can be represented mathematically by a weight matrix:

$$W = \begin{pmatrix} w_{11} \\ w_{21} \\ \cdot \\ \cdot \\ \cdot \\ w_{n1} \end{pmatrix}$$

where w_{ij} represents the influence of concept C_i on concept C_j .

Basically:

$W_{ij} > 0$, an increase or decrease in C_i causes the same result in concept C_j

$W_{ij} < 0$, an increase or decrease in C_i causes the opposite result in C_j

$W_{ij} = 0$, there is no interaction or any influence between concepts C_i and C_j

The magnitude of W_{ij} represents the approximate strength of the influence between two nodes. It usually is determined by a knowledgeable expert or by a mathematical process, observing the process. FCMs therefore differ from ordinary correlation matrices. A correlation matrix is generally symmetric and describes statistical association, whereas an FCM is directional: the causal relationship from X to Y can differ from the relationship from Y to X. This is one of the reasons that researchers confuse the scientific meaning of causality with that of correlation. But taking the correlation coefficient as causality values arrive at the wrong answers. Let us remember that correlation does not imply causality while the reverse is true. Kosko's original formulation [21] described FCMs as fuzzy graph structures for representing causal reasoning, with fuzzy relationships used to describe degrees of causality among concepts.

Mathematics of fuzzy cognitive maps

To understand better this paper, it is needed to know some mathematics of the FCM. From (Figure 2) and the basics of FCM from section 5.2 the mathematics of the FCM theory has been developed extensively [20], [25]. The nodes of the FCM graph stand for concepts that are used to describe, via cause (not correlation) and effect, the relations and behavior of a dynamic system in a simple and symbolic way. They are connected by signed and weighted arcs which represent the causal relationships that exist between the concepts (Figure 2). Each concept, C_i (variable), is characterized by a number that represents its value and is calculated through the transformation of a fuzzy value to the interval [0,1]. The values of the interconnections between the concepts, weights W_{ij} , are initially linguistically defined by experts and then transformed into values which belong to the interval [-1- +1], through a specially designed algorithm, Groumpos [20]. With this algorithm, FCMs embody the

accumulated knowledge and experience from experts who know how the system behaves in different circumstances. The degree of influence between two concepts is indicated by the absolute value of W_{ij} .

During the simulation, the value of each concept (A_i) is

calculated using the following rule:

$$A_i[k+1] = f \left(k_2 A_i[k] + k_1 \sum_{j=1, j \neq i}^N A_j[k] W_{ji} \right) \quad (1)$$

Where:

$A_i(k+1)$: the value of the concept C_i at the iteration step $k+1$

$A_i(k)$: the value of the concept C_i at the iteration step k

W_{ji} : the weight of interconnection from concept C_j to concept C_i

a) k represents the iteration step

b) k_1 controls the contribution of interconnected concepts;

c) k_2 controls the contribution of the previous value of the concept;

d) N is the number of concepts and

e) f is the sigmoid function given by the following equation:

$$f(x) = \frac{1}{1 + B^{-\lambda x}} \quad (2)$$

where $\lambda > 0$ determines the steepness of function f .

The values of the concepts are calculated using Eq. 1. This iterative process ends when a steady state is achieved; for well-defined problems (especially in medical problems) the concepts' values converge always to a single value [29,20,30,27,31].

The FCM approach is based on expert's knowledge for the construction of matrix W_{ij} . This experience is not always reliable though. That is the reason why the weights need to be trained by a learning algorithm [29]. Several learning principles originally developed for ANNs have been applied to FCM modelling. These approaches were based on the concept of Hebbian learning. More specifically Nonlinear Hebbian learning method has been used in many applications. In this learning algorithm the nodes are triggered simultaneously and interact in the same iteration step k with their values to be updated through a process of interaction. This upgrade given by equation (1) describes clearly the dynamic behavior of an FCM. These approaches are given and explained in more detail in [29].

Dynamic behavior of FCMs

One important feature of FCMs is their capacity to represent dynamic systems. Changes in one concept can propagate through the network and affect other concepts. Feedback loops can also be represented. Full details of FCM theories and methods can be found in [20,24] and other papers by the author of this paper [25,26].

For example:

Inflammation $\rightarrow$ oxidative stress $\rightarrow$ cellular damage $\rightarrow$ inflammation

represents a positive feedback structure in which an increase in one process contributes to a sequence of changes that eventually reinforces the original process. This characteristic is particularly relevant to biology because many biological systems are not simple linear chains. Gene-regulatory networks, signaling

pathways, immune responses, metabolic systems, and disease processes contain feedback, redundancy, nonlinear interactions, and emergent behavior. Reviews have documented the use of FCMs across medicine and biological sciences, including decision-support and complex-system modeling [13,18,24,29,32] (Figure 3).



A SIMPLER FCM

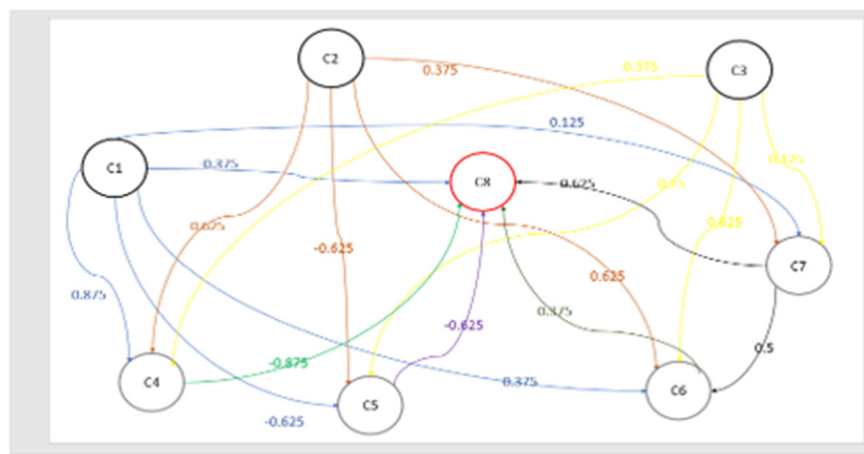


Figure 3: Graphical representation of fuzzy cognitive maps with nodes and weights.

Why should FCMs be Relevant to Biostatistics and Bioinformatics?

The primary advantage of FCMs in biomedical research is their interpretability. A conventional machine-learning model may identify a predictive relationship without making the underlying causal biological assumptions obvious. An FCM, in contrast, produces a visual and mathematical representation of the concepts and relationships being modeled taking into consideration the causality between the concepts.

Consider a hypothetical study examining cardiovascular disease:

Genetic predisposition

- LDL cholesterol
- Inflammation
- Smoking

- Medication
- Blood pressure
- Atherosclerosis
- Cardiovascular events

A conventional statistical analysis might report associations between several of these variables and cardiovascular outcomes. An FCM can place the same variables into a structured network and explicitly distinguish hypothesized directions of influence by analyzing any combination of the variables at the same time in one single recursive equation.

For example: Smoking → inflammation → atherosclerosis → cardiovascular events while simultaneously representing:

Medication → LDL cholesterol → atherosclerosis → cardiovascular events and possible confounding variables such as age, sex, or genetic predisposition.

The graphical representation does not prove these relationships. Instead, it makes the researcher's assumptions inspectable. This is especially useful when biomedical knowledge comes from multiple sources. FCMs can incorporate expert knowledge, literature-derived relationships, and empirical data. Research on participatory fuzzy cognitive mapping has demonstrated the use of FCMs to combine different forms of knowledge, including stakeholder knowledge, systematic reviews, and survey information.

Using FCMs to Distinguish Correlation from Causality

A three-level evidence framework

An important contribution of FCMs can be conceptualized through three levels of evidence to distinguish between Correlation and Causation [8].

Level 1: Statistical association: A relationship is observed in data.

$$X \leftrightarrow Y$$

The direction of the relationship is unknown.

Level 2: Causal hypothesis: Existing biological knowledge or theory suggests that:

$$X \rightarrow Y$$

An FCM can make this hypothesis explicit.

Level 3: Causal evidence: Experimental or formally identified causal analysis provides evidence that manipulating X changes Y, under specified assumptions.

$$do(X=x) \rightarrow Y$$

The critical point is that moving from Level 1 to Level 2 requires reasoning, while moving from Level 2 to Level 3 requires causal identification or experimental evidence. FCMs are particularly valuable at Level 2 because they provide an explicit representation of causal hypotheses.

Example: Gene expression and disease

Suppose a bioinformatics study identifies a correlation between the expression of gene G and disease severity D:

$$\text{corr}(G,D)=0.72$$

The correlation is strong, but several explanations remain possible.

An FCM could initially represent:

$$G \rightarrow D$$

if prior biological evidence suggests that gene G contributes to disease.

However, researchers might also construct alternative models:

$$D \rightarrow G$$

or

$$G \leftarrow Z \rightarrow D$$

where Z is an unmeasured or omitted biological factor.

This process forces researchers to consider competing causal explanations rather than automatically interpreting the statistical association as a causal pathway. This is easily done with FCM.

Integrating FCMs with Biostatistics Data

A useful research workflow can combine data-driven analysis with FCM-based causal reasoning. For the first time using FCM a simple workflow is given below step by step.

Step 1: Data collection

Relevant genomic, transcriptomic, clinical, environmental, and demographic variables are collected.

Step 2: Statistical association analysis

Researchers perform appropriate statistical analyses, such as:

- correlation analysis;
- regression;
- differential expression analysis;
- survival analysis;
- clustering;
- network analysis;
- association testing.

These methods identify candidate relationships but do not automatically establish causality.

Step 3: Candidate relationship selection

Relationships that are statistically significant, biologically plausible, or supported by previous research are selected for further analysis.

Step 4: FCM construction

Variables become nodes, while hypothesized causal relationships become directed edges. Each relationship receives a weight reflecting its expected direction and approximate magnitude.

Step 5: Expert and literature validation

The proposed map is compared with established biological knowledge. Experts can identify implausible edges, missing variables, feedback relationships, and alternative causal explanations.

Step 6: Formal causal analysis

Candidate relationships identified through the FCM are evaluated using appropriate causal methods.

Depending on the research question, these may include:

- Randomized controlled experiments;

- b) Longitudinal studies;
- c) Causal directed acyclic graphs;
- d) Propensity-score methods;
- e) Instrumental-variable approaches;
- f) Mendelian randomization;
- g) Mediation analysis;
- h) Structural equation models;
- i) Counterfactual analysis.

Step 7: Intervention simulation

Once sufficiently justified, the FCM can be used to explore hypothetical interventions and system responses.

The resulting predictions should be treated as model-based scenarios rather than automatically interpreted as experimentally established causal effects.

FCMs and Formal Causal Inference

A major methodological issue is the distinction between an FCM representation of causality and a formal causal model. FCMs are designed to represent causal reasoning, but their edges do not automatically possess the same identification properties as edges in a formally specified causal graph. Formal causal inference asks whether an effect can be identified from available data under explicit assumptions. For example, if researchers want to estimate the causal effect of treatment T on outcome Y, they must consider confounding variables, selection mechanisms, measurement processes, and the assumptions required to identify the intervention effect. Pearl's causal framework emphasizes that causal conclusions depend on assumptions about the underlying data-generating process and cannot be obtained simply from statistical associations.

FCMs can therefore serve as an intermediate layer

Observed data → statistical associations → FCM causal hypotheses → formal causal analysis → experimentally or statistically supported causal conclusions.

This makes FCMs particularly useful for communication and hypothesis generation, especially in medical problems.

Advantages of FCMs in Biomedical Research

Here only a few of the advantages of FCM in Biomedical Research are provided. More can be found on references by this author like in [24].

Interpretability

FCMs provide an intuitive graphical representation of complex systems. Researchers can directly inspect which variables are believed to influence others.

Representation of uncertainty

Biological relationships are rarely perfectly known. Fuzzy weights allow relationships to be represented as degrees of

influence rather than rigid binary statements.

Integration of heterogeneous knowledge

FCMs can incorporate information from statistical datasets, domain experts, published research, and other sources. This is especially important in systems biology, where no single dataset fully describes a biological system. It is also very useful in difficult medical problem which have a similar situation.

Representation of feedback

Many biological processes and difficult problems involve feedback loops. FCMs can represent these loops more naturally than simple linear statistical models, [33].

Explainable computational modeling

Because the nodes and edges correspond to recognizable biological concepts, FCMs can provide a more interpretable alternative or complement to some black-box machine-learning approaches. Systematic reviews [16-18,24,28,32,34] have identified FCMs as widely used for modeling complex causal systems and have documented numerous methodological extensions intended to improve their handling of complex relationships. Something that other probabilistic and/or even AI cannot accomplish at least at this moment. AI might accomplish it in the long future.

Summary and Future Research

Summary

It is important to emphasize that this paper constitutes a conceptual and theoretical review rather than an empirical investigation; no novel experimental datasets or clinical prediction models are presented. The core thesis is that Fuzzy Cognitive Maps (FCMs) offer a specific, highly useful utility within biomedical research, particularly at the intersection of biostatistics and bioinformatics. Traditional statistical approaches are profoundly effective for identifying associations, whereas formal causal inference offers structural frameworks to evaluate interventions under explicit assumptions. FCMs occupy a compelling intermediate territory, empowering researchers to systematically construct, visualize, critique, and communicate hypotheses about variable interactions. This capability is vital in bioinformatics, where biological systems are inherently multidimensional. Faced with the challenge of simultaneously assessing genes, proteins, pathways, environmental exposures, and clinical outcomes, mapping these dynamics as an FCM network renders causal assumptions far more transparent than conventional outputs like regression coefficients [35,36].

A key conclusion of this paper is that distinguishing correlation from causality remains a fundamental challenge in biostatistics and bioinformatics, particularly as biomedical datasets become increasingly complex. FCMs provide a promising framework for addressing part of this challenge by representing biological concepts, directional relationships, uncertain influences, and feedback structures in an interpretable form. They can transform implicit causal assumptions into explicit, inspectable models. However, an

FCM should not be considered evidence of causality simply because it contains directed edges. Its primary contribution is to clarify and communicate causal hypotheses rather than independently establish causal effects. Formal causal inference, appropriate study design, statistical identification, and experimental validation remain essential for drawing strong causal conclusions.

The contribution of FCMs can therefore be summarized as follows: correlation identifies relationships; FCMs make causal hypotheses explicit; and causal inference evaluates whether those hypotheses are justified. Used in this way, FCMs can provide a valuable bridge between statistical analysis, biological knowledge, and rigorous causal reasoning.

Future Research Directions

The most promising direction is an integrated framework in which statistical association analysis identifies candidate relationships. FCMs organize and expose causal hypotheses, and formal causal methods evaluate whether these hypotheses are supported by available evidence. Such an approach could improve the interpretability of biomedical models while reducing the risk of confusing statistical association with biological causation [37-40].

Several research directions could strengthen the integration of FCMs with Biostatistics and Bioinformatics:

- a) **Integration of statistical evidence into FCM weights:** Future research should develop standardized methods for deriving edge weights from statistical evidence, incorporating effect sizes, confidence intervals, longitudinal data, and experimental findings alongside expert knowledge.
- b) **Integration with causal graphical models:** Combining FCMs with causal graphical models could preserve the interpretability and fuzzy reasoning of FCMs while incorporating the identification principles of formal causal inference.
- c) **Development of temporal FCMs:** Time-dependent FCMs could be particularly useful for longitudinal biomedical data, including disease progression, treatment response, and gene regulation.
- d) **Uncertainty and sensitivity analysis:** Researchers should develop methods to quantify uncertainty and assess how sensitive model conclusions are to changes in edge weights. Models producing substantially different conclusions from small weight changes should be interpreted cautiously.
- e) **Benchmarking and validation:** Future studies should compare FCM-based approaches with established causal methods using realistic biomedical datasets and controlled simulations. Such comparisons could clarify where FCMs provide genuine advantages and where traditional causal methods remain more appropriate.

Some visionary and wise thoughts

Before concluding this paper, I feel compelled to share some of my deepest concerns and reflections regarding the present and

future life on our planet. Today, humanity faces an unprecedented range of complex and interconnected challenges. These include issues related to energy and the environment, ecology, business and economics, global stability, and the sustainability of the systems upon which modern civilization depends. Whether environmental, political, technological, social, or economic, many of these systems are proving increasingly fragile, nonviable, or unsustainable, placing humanity at considerable risk. At the same time, we are confronted with extraordinary scientific and technological challenges and opportunities, including Artificial Intelligence, Industry 4.0, intelligent machines, cyber-physical systems, personal safety and security, global poverty, food security, the future of global finance and governance, regional and international conflicts, international stability, and the rapidly evolving Internet of Things (IoT) (World Economic Forum, 2017).

It is becoming increasingly evident that humanity stands at a dangerous crossroads. Unless we reconsider not only the scope of our technological development but also the motives and values that guide it, we may approach a threshold of unimaginable chaos, calamity, and irreversible damage to our planet. This raises a fundamental question: Are there viable solutions to the many problems that have emerged from the rapid expansion of technology and the seemingly unlimited demands that accom. The ultimate resolution may depend on the prudence with which scientific pany knowledge is utilized. Plato articulated this foundational principle over 2,500 years ago: All science, separated from justice and the other virtues, appears to be cunning, not wisdom. (Plato, Menexenos, 246e–247b). Translated simply: Knowledge detached from justice and virtue constitutes cunning rather than true wisdom. This profound assertion remains highly relevant to contemporary discourse.

Epistemological advancement is intrinsically neither virtuous nor beneficial. Possessing extensive knowledge poses significant risks when localized within cohorts that intent to deploy it detrimentally against humanity, or when it is exercised without regard for its broader systemic implications. Knowledge only transmutes into wisdom when guided by, and implemented through, the frameworks of justice, ethical responsibility, and moral virtue. Consequently, technological progression absent a corresponding evolution in humanitarian values creates a profound imbalance—akin to walking with one leg or seeing with only one eye—thereby compromising collective security and future progress. Human society cannot sustain technological acceleration while remaining ethically and socially stagnant.

Paradoxically, despite a global consensus prioritizing peace, prosperity, and sustainable development, contemporary institutions frequently fail to recognize that the resolutions to today's most critical challenges reside within human agency. Whether operating as autonomous citizens, policymakers, or members of the scientific community, it is imperative to cultivate critical inquiry, execute ethical responsibility, and, as prescribed by classical philosophers, engage in genuine and meaningful philosophical praxis. What we need are anthropocentric solutions that place human dignity, well-being, responsibility, and the common good at the center of

scientific and technological progress. In this respect, the famous declaration attributed to the Greek philosopher Protagoras remains deeply relevant:

or in simple English:

“Man is the measure of all things.”

Today, we possess extraordinary scientific disciplines and powerful analytical tools, including biostatistics, bioinformatics, correlation analysis, and causal inference. When used wisely, responsibly, and ethically, these disciplines can contribute to virtually every field of science and help us develop evidence-based solutions to some of the world's most difficult challenges.

Yet, scientific knowledge alone is not enough. The future of humanity will depend not simply on how much we know, how intelligent our machines become, or how advanced our technologies are, but on how wisely we choose to use them. Finally, perhaps the most important thought with which I would like to conclude is this: we do not need to become smarter or more intelligent; we need to become wiser.

Conclusion

This paper considers, for the first time, whether and how the new emerging scientific discipline of Fuzzy Cognitive Maps (FCMs) can be useful for addressing challenging research problems in medicine, particularly in Biostatistics and Bioinformatics. FCMs provide an alternative approach to analyzing complex and large datasets. Rather than relying exclusively on statistical methods, they can incorporate expert knowledge and experience, offering a way of representing complex systems in a manner that reflects human reasoning. Importantly, FCMs can transform implicit causal assumptions into explicit and inspectable models and may help researchers explore whether proposed causal relationships are consistent with available data. A key conclusion of this paper is that distinguishing using FCMs, correlation from causality remains a fundamental challenge in Biostatistics and Bioinformatics, particularly as biomedical datasets become increasingly complex. FCMs provide a promising framework for addressing part of this challenge by representing biological concepts, directional relationships, uncertain influences, and feedback structures in an interpretable form. This distinction is particularly important in clinical research especially for young researchers. For example, a model may suggest that increasing a biomarker would improve a clinical outcome. However, an intervention based only on this prediction based in correlation could be harmful if the biomarker is merely a downstream indicator rather than a causal factor.

The value of FCMs is therefore not that they automatically transform correlations into causal relationships, but that they make causal assumptions explicit. Researchers must specify which concepts influence others, in what direction, and with approximately what strength. This can reveal assumptions that might otherwise remain hidden within statistical models or computational pipelines.

The strongest application of FCMs is consequently not to replace randomized experiments or formal causal-inference methods,

but to help researchers identify important causal questions and determine which assumptions require further investigation. Recent literature highlights the need for greater methodological transparency and stronger integration between fuzzy modeling and formal causal frameworks. Introducing principles from Cybernetics into medical research, particularly Biostatistics and Bioinformatics, may provide valuable tools for addressing complex problems in health sciences. As discussed in [2], interactive knowledge discovery and data mining in biomedical informatics can help address difficult research challenges by combining advanced theories and methods from different scientific disciplines. The new theoretical discipline of Fuzzy Cognitive Maps (FCM) can be of some help in addressing several difficult but challenging issues of Biostatistics and Bioinformatics.

References

1. Shi Y (2022) *Advances in big data analytics*. Springer Nature, Singapore.
2. Holzinger A, Jurisica I (2014) *Interactive knowledge discovery and data mining in biomedical informatics*. State-of-the-Art and Future Challenges, Lecture Notes in Computer Science (LNCS) Volume 8401, Springer Publishers, UK.
3. Bland M (2015) *An introduction to medical statistics*. (4th edn.), Oxford University Press, UK.
4. Ashby D (2006) Bayesian statistics in medicine: A 25-year review" *Statist Med* 25(21): 3589-3631.
5. Rohrer JM (2018) Thinking clearly about correlations and causation: Graphical causal models for observational data. *International J Advances in Methods and Practices in Psychological Science* 1(1): 27-42.
6. Shakeri R, Hossein A, Farshid F, Man YL, Banafsheh Z (2025) Comparative analysis of correlation and causality inference in water quality problems with emphasis on TDS Karkheh River in Iran. *Sci Rep* 15: 2798.
7. Madadzadeh F, Bahariniya S (2026) Association, correlation, and causation: Three important and misleading terms in research. *The Open Public Health Journal* 19(3).
8. Zaniletti I, Larson DR, Lewallen DG, Berry DJ, Kremers HM, et al. (2023) How to distinguish correlation from causation in orthopedic research. *J. Arthroplasty* 38(4): 634-637.
9. Rotari M, Kulahci M (2025) Correlation to causality. *Quality Engineering* 37(1): 162-172.
10. Pearl J (2010) An introduction to causal inference. *The International Journal of Biostatistics* 6(2): 7.
11. Pearl J (2010) Causal inference. *Proceedings of the Workshop on Causality: Objectives and Assessment at NIPS 2008, Proceedings of Machine Learning Research*, pp.39-58.
12. Yang HJ, Lee HS (2025) Common statistical methods used in medical research. *Kosin Medical Journal* 40(1): 21-30.
13. Batko K, Ślęzak A (2022) The use of big data analytics in healthcare. *Journal of Big Data* 9(1): 3.
14. Holland PW (1986) Statistics and causal inference. *Journal of the American Statistical Association* 81(396): 945-960.
15. Prakash K, Ranjit KP, Himadri SR, Md Yeasin, Ajit, et al. (2024) Big data analysis in computational biology and bioinformatics. *Methods Mol Biol* 2719: 181-197.
16. Schuerkamp R, Giabbanelli PJ (2023) Extensions of fuzzy cognitive maps: A systematic review. *ACM Computing Surveys* 56(2): 1-36.
17. Sharma A, Tselykh A, Podoplelova E, Tselykh A (2022) Knowledge-oriented methodologies for causal inference relations using fuzzy cognitive maps: A systematic review. *Computers & Industrial*

Engineering 171.

18. Jamett J, Melendez P, Collao-Ferrada X, Cordero-Torres K, Veloz A (2026) Fuzzy logic approaches for causal inference in health care: Systematic review. *JMIR AI* 5: e83425.
19. Zakaria A (2026) A human-centric fuzzy decision support system for medical diagnosis using fuzzy cognitive maps. *Sci Rep* 16: 15336.
20. Groumpos PP (2010) Fuzzy cognitive maps: Basic theories and their application to complex systems. In: Glykas M (Ed.), *Fuzzy cognitive maps: Advances in theory, methodologies, tools and applications*. Springer Publishers UK, pp. 1-22.
21. Kosko B (1986) Fuzzy cognitive maps. *International Journal of Man-Machine Studies* 24(1): 65-75.
22. Kosko B (1968) Fuzzy knowledge combination. *International Journal of Intelligent Systems* 1(4): 293-320.
23. Axelrod RM (1976) *Structure of decision. The cognitive maps of political elites*. (1st edn), Princeton University Press, USA, pp. 1-422.
24. Groumpos PP (2017) Why model complex dynamic systems using fuzzy cognitive maps? *Robot Autom Eng Journal* 1(3): 57-69.
25. Groumpos PP, Stylios CD (2000) Modelling supervisory control systems using Fuzzy cognitive maps. *Chaos Solutions and Fractals* 11(1): 329-336.
26. Anninou AP, Groumpos PP, Polychronopoulos P (2013) Modeling health diseases using competitive fuzzy cognitive maps. *Artificial Intelligence Applications and Innovation*, pp. 88-95.
27. Shanti KL, Gogada JSG, Satyanarayana SS, Gogada R (2021) Bioinformatic analysis of metal transportomes from *Mycobacteria* Sp. *Open J of Biostatistics and Bioinformatics* 5(1): 14-22.
28. Amirkhani A, Papageorgiou EI, Mohseni A, Mosavi MR (2017) A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications. *Computer Methods and Programs in Biomedicine* 142: 129-145.
29. Papageorgiou EI, Stylios CD, Groumpos PP (2004) Active hebbian learning algorithm to train fuzzy cognitive maps. *Intern J of Approximate Reasoning* 37(3): 219-249.
30. Papageorgiou EI (2011) A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. *Applied Soft Computing* 11(1): 500-513.
31. Groumpos PP, Anninou AP (2012) A theoretical mathematical modeling of Parkinson's disease using fuzzy cognitive maps. *Bioinformatics & Bioengineering (BIBE), IEEE 12th International Conference*, pp. 677-682.
32. Krimpas GA, Krimpas NA, Groumpos PP (2023) An overview of making decisions on medical problems using fuzzy cognitive maps. *14th International Conference on Information, Intelligence, Systems & Applications (IISA), Greece*, pp. 1-6.
33. Apostolopoulos ID, Groumpos PP (2023) Fuzzy cognitive maps: Their role in explainable artificial intelligence. *Applied Sciences* 13(6): 3412.
34. Aguilar J (2005) A survey about fuzzy cognitive maps papers. *Inter Journal of Computational Cognition* 3(2): 27-33.
35. Eticha AB (2020) Forecast number of new cases of Corona Virus Disease (COVID-19) in Ethiopia, using the case-based autoregressive integrated moving average model. *Open J Bioinform Biostat* 4(1): 17-22.
36. Papandrianos N, Papageorgiou E, Anagnostis A, Feleki A (2020) A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans. *J Applied Sciences* 10(3): 997.
37. Groumpos PP (2021) Modelling COVID-19 using Fuzzy Cognitive Maps (FCM). *EAI Endorsed Transactions on Bioengineering and Bioinformatics* 1(2): 168728.
38. Holzinger A, Jurisica I (2014) Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. *Interactive knowledge discovery and data mining in biomedical informatics*, pp. 1-18.
39. Groumpos PP (2018) Making the world a better place to live through wisdom and philosophy. *IFAC-Papers Online* 51(30): 744-749.
40. Carvalho JP, Tome J (2002) Issues on the stability of fuzzy cognitive maps and rule-based fuzzy cognitive maps. In *Proceedings of the 21st International Conference of the North American Fuzzy Information Processing Society, NAFIPS2002, USA*.