2022

Investigation of the relation between Human's brain object recognition and its working memory using advanced EEG signal analysis and ML techniques

By Cyrus Kalantarpour

Investigation of the relation between Human’s brain object recognition and its working memory using advanced EEG signal analysis and ML techniques

Contents

I believe that Neuroscience and Artificial Intelligence have pushed each other forward. The concept of artificial neural networks (ANNs), which is inspired by the concept of biological neural networks, has undoubtedly altered AI history. When we look at the milestones of ANNs like convolution, long-term short memory, attention, learning algorithms, and so on, we can see how bio-thoughts influenced them. For this reason, I would like to use analysis the role of brain's different parts communications in object recognition task.

In the following I tried to propose a few questions and answers to provide the reader with a better perspective.

What is the purpose of the work?

• A1: Discovering how humans brain memory interact with vision cortex.

• Q2: Which discipline and which themes are the choice of research topic?

• A2: Furrier Transform, Wavelet transform, Correlation analysis, Cross-frequency coherence analysis. Artificial Intelligence (AI), Machin-Learning (ML), Neural networks (NNs).

• Q3: Why the research topic is feasible?

• A3: There are lots of public EEG datasets and all the above methods are implemented in Python language.

• Q4: How do I compare my work with the other similar works?

• A4: Our comparison is quantitative, and the classification of processed EEGs will be compared by the accuracy measure.

• Q5: What is the research milestone(s)?

• A5: Put a step forward to understanding how different parts of the brain are communicating. It is worth noting that this information is very useful to improve computer vision neural networks.

• Q6: Why I decided to investigate human brain data?

• A6: I think Neuroscience and Artificial intelligence have pushed forward each other. It is obvious that history of AI has been revolutionized by the concept of artificial neural networks (ANNs) which is inspired by the concept of biological neural networks. If we look at the milestones of the ANNs such as, convolution, Long-term short memories, Attentions, learning algorithms and… we can see they have inspired by bio-thoughts.

• Q7: What is the type of my research?

• A7: It is an applied research.

• Q8: Which type of the methodology are I using in my research?
• A8: Quantitative.
• Q9: What data am I using in my research?
• A9: Public EEG datasets.
• Q10: What are the possible consequences of my research?
• A10: Improving the computer vision.
• Q11: What will be my future works?
• A11: Implementation of the inspired neural architecture.

Introduction

We investigate the human brain electrical activity during a extensive brain stimulation during an sensory- cognitional task. More specifically, we investigate the human brains EEG signal while the clients are controlling a flight simulator. The intuition behind choosing of the aforementioned dataset is the ability of extracting the communication features of different human brain regions. That is because, during controlling of a flight simulator, most of the brains regions such as visual processor, audio processor, working memory, sensory motor (muscle control) and … should have an synchronized intercommunications and cross communications. As a result, this investigation may inform us how the human brain is able to use its different neural circuits to execute a complex task. It is worth noting that one of the main steps toward the Artificial General Intelligence is incorporation and integration of different types of neural architectures to overcome human- level task such as causal reasoning.

EEG

The electroencephalogram (EEG) is a record of the oscillations of brain electric potential recorded from electrodes on the human scalp. Consider the following experiment. Place a pair of electrodes on someone's scalp and feed the unprocessed EEG signal to a computer display in an isolated location. Independently monitor the subject's state of consciousness and provide both this information and the EEG signal to an external observer. Even a naive observer, unfamiliar with EEG, will recognize that the voltage record during deep sleep has larger amplitudes and contains much more low-frequency content. In addition, the eyes closed waking) alpha state will be revealed as a widespread, near-sinusoidal oscillation repeating about 10 times per second (10 Hz). More sophisticated monitoring allows for accurate identification of distinct sleep stages, depth of anesthesia, seizures, and other neurological disorders. Other methods reveal robust EEG correlations with cognitive processes associated with mental calculations, working memory, and selective attention.

Ten/20 coordinating system

The Ten/20 system is a measurement system that divides the human scalp into 19 sections and measures it. Each electrode of the EEG recording device will be positioned on one of these 19 locations in this arrangement. This system is in place. Pre-frontal (Fp), frontal (F), temporal (T), parietal (P), occipital (O), and central (C) are the names of the brain regions (C). The idea behind this brain partitioning is that distinct brain areas (Lobes) are important in different cognitive tasks. For example, visual processors are in the occipital lobes, working memory is in the temporal lobes, and auditory lobes are in the auditory lobe....

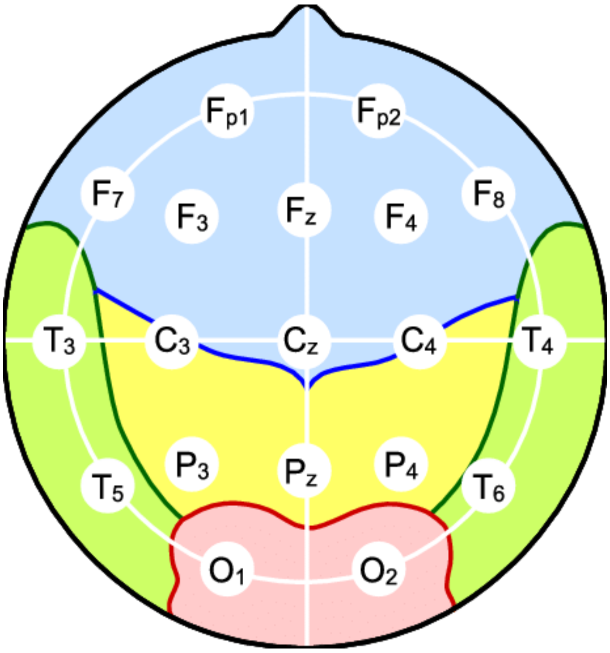

Figure 1 Demonstration of the 10/20 measurement system. Each white circle shows an EEG electrode placement. The colored areas show different brain's lobe.

Artifacts

Electroencephalogram (EEG) technic helps healthcare professionals and researchers to detect and monitor brain activities and behavior. However, while monitoring brain activities, the artifacts which are non-brain activities such as blinking can interfere with the EEG procedure. Qi et al (2021) categorized common artefact components as blinking the eyes, horizontal or vertical eye movements, and generic discontinuities. More artifacts are categorized by Muscle movements artifacts such as: Respiration artifact –artifacts generated by talking and tongue movements and sweating.

Many algorithms and methods are developped for removing artefacts. However, there is no specific algorithm capable of removing all types of artifacts (Jiang, Bian, and Tian 2019).

Those contaminated EEG signal which prevent accurate EEG interpretation. There are two types of artifacts: – Nonphysiologic artifact (not from the patient) – Physiologic artifact (from the patient).

Regarding the non physiologic, possible sources could be: EEG device (Electrodes, Headbox, Amplifier, Cable, Environment and...)

Artifact rejection methods

There are two main technics for Artifact rejection: 1)- Neural Networks such as [THE (Svoboda et al. 2016)] 2- statiscital methods (Nolan, Whelan, and Reilly 2010).

In the neural method usually use a convolution block and train the networks (Svoboda et al. 2016). Compared to the statisitcal method, neural method are more accurate but their accuracy may vary on different test sets[THE CONVOLUTION PAPER??] In the following we see the some statistical artifacts rejection methods.

Fully Automated Statistical Thresholding for EEG artifact Rejection (FASTER), Independent component analysis (ICA) and wavlet ICA.

Compare to the ICA, FASTER is faster when applied to high-density EEG data. wavelet ICA

Brains different regions' cognitive function

According to the Johns Hopkins' neurology department.

Frontal lobe. The largest lobe of the brain, located in the front of the head, the frontal lobe is involved in personality characteristics, decision-making and movement. Recognition of smell usually involves parts of the frontal lobe. The frontal lobe contains Broca's area, which is associated with speech ability.

Parietal lobe. The middle part of the brain, the parietal lobe helps a person identify objects and understand spatial relationships (where one's body is compared with objects around the person). The parietal lobe is also involved in interpreting pain and touch in the body. The parietal lobe houses Wernicke's area, which helps the brain understand spoken language.

Occipital lobe. The occipital lobe is the back part of the brain that is involved with vision.

Temporal lobe. The sides of the brain, temporal lobes are involved in short-term memory, speech, musical rhythm and some degree of smell recognition.

EEG data cross-correlation analysis

In this phase of this study, we wanted to test whether there are correlations between EEGs segments using cross-correlation technique.

The cross-correlation between two EEG channels measures the level of dependency between their frequency components. More specifically, If the cross-correlation of two EEG channels is 1 at a time $t0$, the EEGs are behaving similarly. That is, the amplitudes for the two channels are equal. Similarly, if the cross-correlation between two EEG channels is 0.8 at time $t1$, then the EEGs are behaving similarly at the $t1$ and the amplitudes of their components are 80% equal. Any negative correlation means the channels frequency components are correlated but their behavior is the opposite. Furthermore, the values around zero means EEG channels are acting independently (Dowdy, Wearden, & Chilko, 2011).

In the following we will illustrate all brain lobes intercommunications and communications in terms of the cross - correlation. Using the cross correlation will able us to discover the brain lobes communication harmonics and consequently their communication method which can help us improv e the existing artificial neural architectures.

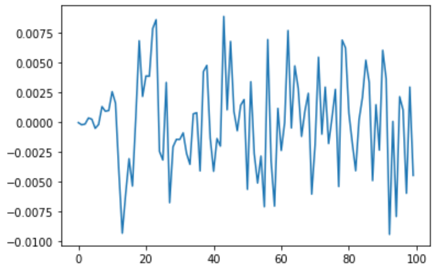

Figure 2:Pre-Frontal and Frontal lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

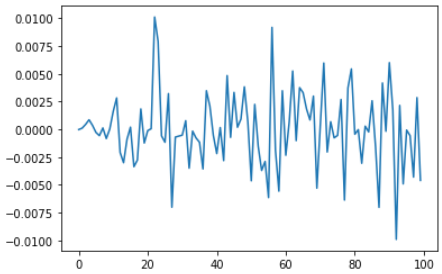

Figure 3 Pre-Frontal and Temporal lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

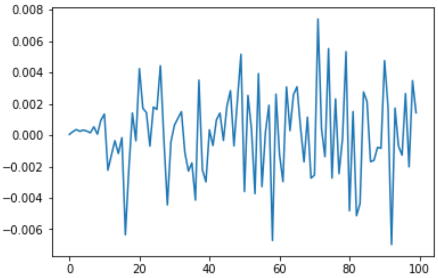

Figure 4 Pre-Frontal and Central lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

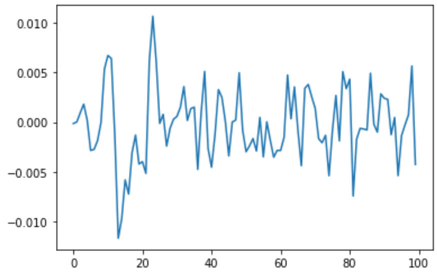

Figure 5 Pre-Frontal and Parietal lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

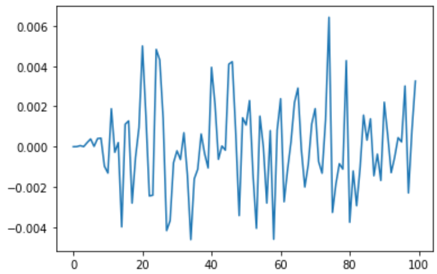

Figure 6 Pre-Frontal and Occipital lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

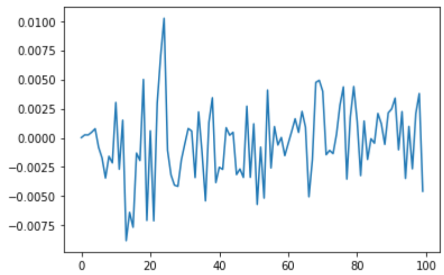

Figure 7 Frontal and Central lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

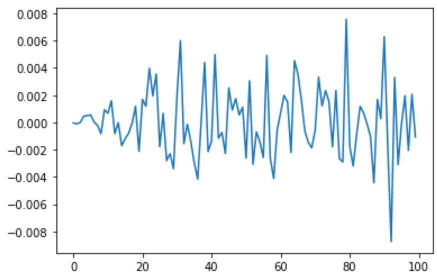

Figure 8 Frontal and Temporal lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

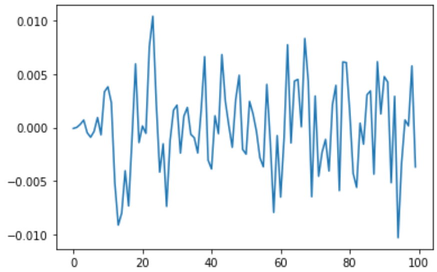

Figure 9 Frontal and Parietal lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

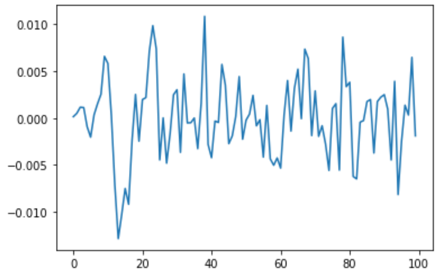

Figure 10 Frontal and Occipital lobe Cross -Correlation. The x axis is frequency and y axis the correlation value.

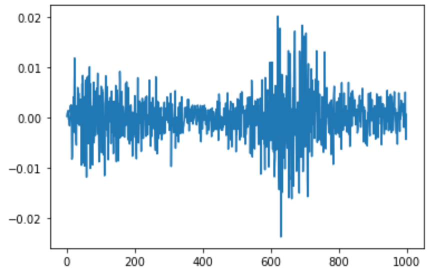

Figure 11 Total Cross correlation of the Pre-frontal and left occipital. The x axis is frequency and y axis the correlation value.

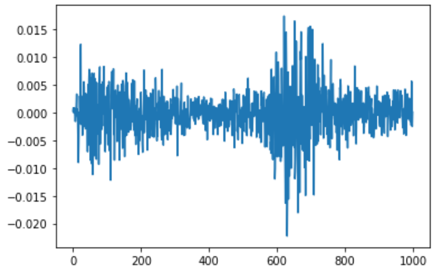

Figure 12 Total Cross correlation of the Pre-frontal and right occipital. The x axis is frequency and y axis the correlation value.

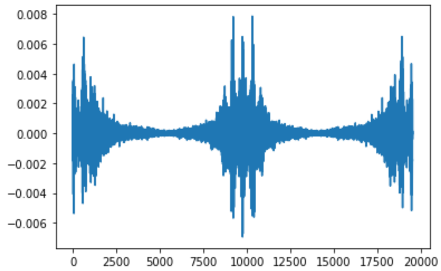

Figure 13 differentiation of Total Cross correlation of the Pre-frontal and left occipital with Pre-frontal and right occipital. The x axis is frequency and y axis the correlation value.

FINAL PHASE

Intro

In the previous phase of my project, I have investigated the 19-channel EEG data of pilots while they are controlling the flight simulator. The reason is the brain has to use its different neural networks including frontal, central and the parietal lobes together. Therefore, using this data we can figure out how the brain can incorporate its different neural networks with different cognitive functions to achieve a complex task such as controlling an airplane.

Final results

In the primary stage of the project I used the correlation technique to figure out how the two different part of the brain with different function can collaborate to do a complex task. But, I did not manage to achieve a meaningful result. Therefore, I tried to use more advanced tools such as Artificial neural networks to obtain a good result. To do so, I used the CNNs and CNN+LSTM (from my proposed architecture in my first publication https://journals.flvc.org/FLAIRS/article/view/128367/130120). Although, the CNN+LSTM architecture basically has spectacular power for feature extraction, again it cannot properly extract the features of the dataset.

The reason is conventional artificial neural networks only can recognize the patterns and correlation while in this study the correlation cannot help. Since we need to know whether the activation of a specific part of the brain is the cause to activation of another part. That is, the correlations are useless since it is possible that two different parts of the brain activates together but those have not any causal relationship. In the following I will give you an example illustrate the difference between correlation and causation.

Example: In the Figure (14) you can see if the laptop runs out of battery it would be a cause to the computer shuts or making the video player not to work. While, the computer shuts would not be a cause of the video player won't work.

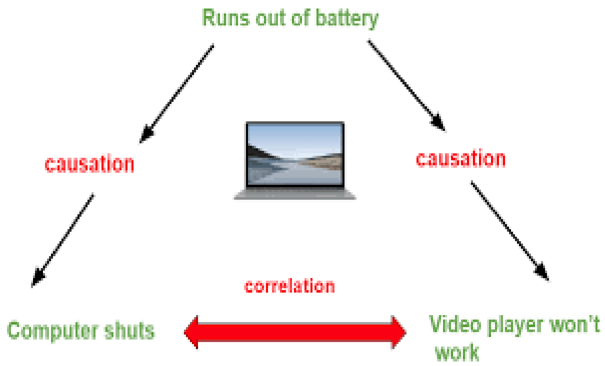

Figure 14:Causation vs Correlation

Therefore, according the the above, I used my novel architecture (fuzzy causal effect variational autoencoder (FCEVAE) proposed in my second publication) to capture the causal relationship of the brain's different part activity.
To become familiar with the mechanism of the FCEVAE, we need to study the following prerequisites.

1. Fuzzy logic
2. Fuzzy inference
3. Variational Autoencoder
4. Causal effect variational autoencoder
5. Fuzzy Causal Effect Variational Autoencoder (My second publication)
6.

Since the above topics from 1 to 4 are too vast I decided not to include any description for them (please use the references). Regarding the 5, You can read the publication from the bellow.

Causal Probabilistic based Variational Autoencoders capable of handling noisy inputs using Fuzzy Logic Rules

Usef Faghihi[1], Cyrus Kalantarpour[1] and Amir Saki[2]

[1] University of Quebec at Trois-Rivières, QC, Canada
[2] Institute for Research in Fundamental Sciences, Iran
usef.faghihi@uqtr.ca, cyrus.kalantarpour@uqtr.ca, amirsaki1369@ipm.ir

Abstract. Researchers and engineers may use inferential logic and/or fuzzy logic to solve real-world causal problems. Inferential logic uses probability theories, while fuzzy logic uses its membership functions and set theories to process uncertainty and fuzziness of the events. To benefit from both logics, some researchers in the past tried to create probabilistic fuzzy logic (PFL). Deep Learning algorithms (DLs) with their incredible achievements such as very high precision results in some specific tasks are at the center of the weak AI. However, DLs fail when it comes to causal reasoning. In order to equip Deep Learning algorithms (DLs) with reasoning capabilities, one solution would be to integrate non-classical logics such as PFL with DLs. In this paper, we will demonstrate the first step toward creating a deep causal probabilistic fuzzy logic architecture capable of reasoning with missing or noisy datasets. To do so, the architecture uses fuzzy theories, probabilistic theories, and deep learning algorithms such as causal effect variational autoencoders.

Keywords: Deep Learning, Probabilistic Fuzzy logic, Causal reasoning, Autoencoders.

1 Introduction

As human beings, we are always in the search for the causes of events around us. For instance, was it the spicy food I had for my lunch that caused my abdominal discomfort? In causal reasoning, one uses previous information about an event or situation to predict its future state. However, discovering the real causes of events is usually difficult.
In order to solve the problem of causality, some researchers use inferential logic, which uses probability theory, while others use fuzzy logic, which outperforms inferential

logic [3]. Probability theories deal with the uncertainty of human knowledge about an event. However, there is no gradient possible with probability theories [3]. Fuzzy logic makes it possible to take into account the vagueness of events. The ideal would be to use both inferential and fuzzy logic theories together and create probabilistic fuzzy logic [3]. In [3], Faghihi et al. used causal fuzzy rules belonging to fuzzy rule sets to find the influences of confounders on other variables. A confounder variable influences both dependent and independent variables causing fake correlations between variables. However, Faghihi et al's model does not have learning capabilities [3]. To equip PFL with learning capabilities, one must integrate them into the DLs [4]. One powerful generative deep learning algorithm that is widely used to deal with different real-world problems is Variational Autoencoders architectures families [2, 5]. In the following, we briefly explain Autoencoders, Variational autoencoders, and Causal Effect Variational Autoencoders [2].

Autoencoder (AE) is a specific type of generative artificial neural network that learns representation for a set of data in an unsupervised manner. An AE [2] has 1) Encoder module (inference network in causality context) which encodes or compresses the input data into a latent space representation (a reduced version of the original data); 2) Decoder module which tries to reconstruct the original input data from the latent encoded space.

Variational Autoencoders (VAEs) are similar to the AEs, except they consider a family of Gaussian distributions while sampling from the input data. VAEs work with both continuous and discrete data. Recently, researchers created Causal Variational Autoencoders (CEVAE) [2], which estimate the individual and average treatment effects (ITE, and ATE respectively) for unobserved confounders using proxy variables which are replacements for confounders [3].

In this paper, we first discuss the related works on causality using different DLs such as variational autoencoder families. In order to extract causal relationships from observational data, we then discuss two architectures that use PFL and causal effect variational autoencoders (CEVAEs) architecture [2] which we call FCEVAE-V1 and V2. The first architecture is called FCEVAE -V1: in this architecture PFL and CEVAE each separately applied to the dataset. That is, we cluster and fuzzify the data set, and then, it is feed-forwarded to the CEVAE architecture (Figure 1). The second architecture is called FCEVAE -V2: we integrated association and causal rules from [3] to the CEVAE loss function (Figure 2). That is, in the second architecture, we equipped CEVAE with a modified loss function that implements causal fuzzy logic rules from [3]. It must be noted that the initial CEVAE architecture developed by Amsterdam lab[1] uses TensorFlow and Edward (deprecated). In this study, we used a CEVAE version equipped with Pyro library[2]. Pyro is faster than Eward.

Finally, we compare the performance of our architecture with similar architectures and discuss the results and limitations of our work.

---

[1] https://github.com/AMLab-Amsterdam/CEVAE

[2] https://github.com/pyro-ppl/pyro

2 Related Works

Recently, learning causal relationships from observational data received lots of attention in the field of Artificial Intelligence [6, 7]. Moreover, some researchers try to address the identifiability[3] issue using neural networks [8]. However, the observational data may contain hidden confounder variables that may not or very difficult to be measured [2].

Take a study in which we are interested in individualized medicine and where we have to figure out the best medication for a patient from observed data [2]. In this example, the socio-economic status of the patient can influence the type of medication the patient has access to and her general health [2]. That is, the socio-economic status is a confounder, and we cannot compute its value [2]. It is worth mentioning that once we can estimate or calculate the confounder's value, another hurdle to overcome is to find which element(s) it influences the most.

Let's suppose we cannot measure the confounder which is the socioeconomic status of the patient. Roughly speaking, there are two main approaches to calculating confounders. The first one is a tree-based approach [9], wherein the authors use Bayesian Additive Regression Trees (BART) [10] to estimate average causal effects for solving causal inference problems such as individual treatment effects (ITE). The second approach uses Directed Acyclic Graphs (DAGs) as a causal structure and a Bayesian approach for reasoning.

TARnet [11] is one of the first architectures that is used for causal inference. It is based on weighting optimizations and using the feed-forward neural networks. However, TARnet is not robust enough to deal with noisy datasets [2, 12]. In 2017, Louizos et al [2] created Causal Effect Variational Autoencoders (CEVAE) which estimates the individual and average treatment effects (ITE and ATE) for unobserved confounders using proxy variables. A confounder variable that can be hidden and/or have missing data, influences both dependent and independent variables, causing fake correlations between variables. The model suggested by the authors in [2] outperformed Tree-based approaches such as BART [2]. However, the model in [2] has problems with processing missing data.

To improve CEVAE, the authors in [12] created Identifiable VAE (iVAE) architecture. This architecture postulates that different model parameters must lead to the different marginal densities for the data. In 2021, the authors suggested Intact VAE [13], an improved version of iVAE. Intact VAE estimates ATE by using a modified version of propensity score (the probability of a subject receiving treatment) and B-score (The conditional distribution for the covariates receiving or not receiving treatment is the same). However, this study ignores computing confounders. As opposed to current DLs which can only process either noisy or missing data, a robust DL needs to be both tolerant to both noisy and missing data with hidden confounders. We will achieve this by integrating Non-Classical Logics such as probabilistic fuzzy logic rules with DLs.

[3] If the true parameters of a statistical model can be learned after observing sufficient number of observations, the model is said to be identifiable. Wikipedia

Faghihi et al [1] argued that in most real-life problems, the communication between nodes is two-way, something DAG does not support. In other words, the mere Bayesian approach to causation cannot answer the following problem: what is the probability that socio-economic status influences the type of medication the patient has access to and her general health, and to what degree?
Probabilistic Fuzzy logic (PFL), on the other hand, excels at reasoning with degrees of certainty and in real-life problems [14]. Importantly, this allows for degrees of dependency and membership. In PFL, Zadeh [14] proposes that a given set of elements always has a degree of membership and fits into an interval between [0,1]. PFL processes three types of uncertainty: randomness, probabilistic uncertainty, and fuzziness.
PFL can both manage the uncertainty of our knowledge (by the use of probabilities) and the vagueness inherent to the world's complexity (by data fuzzification) [14]. PFL has been used to solve many engineering problems, such as security intrusion detection [15, 16] and finding the causes of the events. However, PFL cannot learn by itself and needs experts to define intervals before applying fuzzification [3]. In [3], the authors used more than ten PFL rules to discover the causal relationship between variables from observational data. However, logic can not learn a representation of the data [3]. One solution would be to integrate PFL with Deep Learning algorithms or use them in parallel. In the next section we explain how we used CEVAE architecture [2] with PFLs.

3 FUZZY CEVAE

We designed and implemented two versions of the CEVAE architecture [2] which we call FCEVAE: 1) FCEVAE-V1: in this architecture PFL and CEVAE separately applied to the dataset. That is, we clustered and fuzzified the dataset using PFL and then feed forwarded it to the CEVAE architecture (Figure 1); and 2) FCEVAE-V2: in this architecture, we integrated clustering and causal rules with the CEVAE architecture (Figure 2). That is, in the second architecture, we equipped CEVAE with a modified loss function that implements causal fuzzy logic rules from [3].

3.1 Fuzzy Causal Effect Variational Autoencoder (FCEVAE-V1) First Architecture

To cluster the dataset into "Low", "Average", and "High" clusters, we used the fuzzy c-mean algorithm [1] (Figure 1. A). It is worth mentioning that depending on the problem, one can use more than three clusters if needed.

However, the C-mean clustering only gives us the membership belongingness of the dataset elements to every clusters. Thus, C-mean's output does not include any information about the nature of the dataset. Consequently, we multiplied the clustered data with the original dataset. This gave us a weighted fuzzy representation of the dataset elements describing how well each element belongs to the fuzzy clusters.

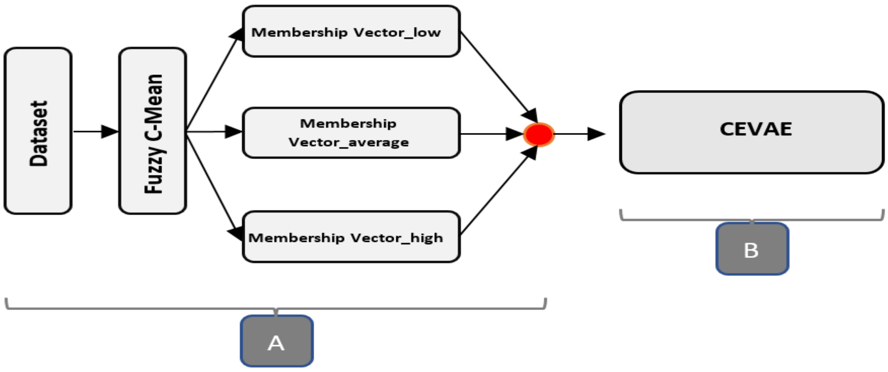

Figure 1:Part A is a membership vector extractor according to the Fuzzy c-mean algorithm [1]. It applies Fuzzy c-mean on the dataset and computes memberships vectors for the elements of the dataset. Then, the red circle, which in our context is a 'switch neuron', selectively multiplies the memberships vectors calculated in the previous step to the original dataset. Part B is the original CEVAE proposed in [2].

To justify our above multiplication, we briefly explain our Simple Probabilistic Fuzzy Logic Theory (SPFL) theory [17], a classical probability theory mainly useful for the problems with fuzzy concepts in their nature. For instance, the following problem could be solved by using the SPFL theory.

Question. Suppose the fuzzy attribute *Large* for the set $X = (1, \ldots, 20)$. In an experiment, what is the probability of randomly selecting 17 from X as Large?"

To answer the question, one can define random variable $\xi_{X,Large}$ so that $\mathbb{P}(\xi_{X,Large} = 17)$. Hence, here the distribution of $\xi_{X,Large}$ matters. Randomly selecting the elements of X comes from the nature of the distribution on X, while selecting as Large comes from a two steps procedure consisting of fuzzifying data by some fuzzy attributes including Large, and then the distribution determining the chance of being selected as $Large$.

Another example follows:

Question. Suppose the fuzzy attribute *Large* for the set $X = (1, \ldots, 20)$. In an experiment, we are given $X = 17$. What is the probability of selecting 17 as Large?

The answer to this question is $\mathbb{P}(x \text{ is } Large)$, and it comes from a distribution. Now, a binary random variable is considered as follows:

$$\xi_{x,Large} = \begin{cases} x & , \mathbb{P}(x \text{ is } Large) \\ 0 & , 1 - \mathbb{P}(x \text{ is } Large) \end{cases}$$

That is, $\mathbb{E}(\xi_{x,Large}) = x\mathbb{P}(x \text{ is } Large)$, and it is interpreted as the quantity of x as being *Large*. Note that, in this paper we use a model with $\mathbb{P}(x \text{ is } Large) = \mu_{Large}(x)$.

To calculate **Fuzzy Average Treatment Effect (FATE)** which will be used in our below FCEVAE (second architecture), we perform as follows: Suppose X, T and Y are the covariate, the treatment and the outcome of an experiment, respectively. Let A be a fuzzy attribute of X. We define the *fuzzy individual treatment effect* of any $X = x$ with respect to A as: $\mathrm{FITE}_A(x) = \mathrm{ITE}\left(\mathbb{E}(\xi_{x,A})\right)$.

It follows that:

$$\mathrm{FITE}_A(x) = \mathbb{E}(Y|X = x\mu_A(x), \mathrm{do}(T = 1)) - \mathbb{E}(Y| = x\mu_A(x), \mathrm{do}(T = 0)).$$

Now, we define FATE of X with respect to A as $\mathrm{AFTE}_A(X) = \mathbb{E}_X(\mathrm{FITE}_A(X))$.

Going back to the FCEVAE-V1 architecture (Figure 1), by multiplying the clustered data with the original dataset, we obtain a weighted fuzzy representation of the dataset elements describing how well each element belongs to fuzzy clusters. As a result, FCEVAE-V1 produces three different average treatment effects values [18] each describing the fuzzy average treatment effect corresponding to the clusters such as “low”, “average”, and “high”. Table 1 shows that FCEVAE-V1 outperforms Microsoft’s DoWhy[4] project that implements Pearl’s causal architecture [19] on Infant Health and Development Program (IHDP) [9] dataset. IHDP dataset contains information about the effect of specialists’ home visits on premature infants’ cognitive test scores [6]. In addition, the average number of ATEs obtained by FCEVAE-V1 is 4.006. This value is closer to the real IHDP’s ATE of 4.021 [19]. It is worth noting that in the DoWhy project, the ATE value for the IHDP dataset was calculated by subtracting the mean of the treated and controlled groups.

Cluster	**Low**	**Average**	**High**
FCEVAE-v1	3.812	4.015	4.192
Microsoft DoWhy	3.928		

Table 1: Comparison of FCEVAE-v1 and DoWhy.

However, our first architecture has two flaws: 1) similar to the original CEVAE architecture [2], to select the treatment and outcome columns, the architecture needs a human expert. However, in a real-world problem, humans may have no idea about the Treatment and Outcome columns; 2) Because we fuzzify the dataset before feeding it to the CEVAE architecture, it cannot tolerate noisy data. We fixed the first architecture’s flaws in our second architecture.

Unlike the first architecture that uses fuzzy weighted versions of datasets to create fuzzy-probabilistic-based CEVAE architecture (without using fuzzy rules), the second architecture incorporates fuzzy causal rules from [3, 20] in its loss function. This helps the CEVAE architecture discover the causal relationships between the dataset's elements.

[4] https://microsoft.github.io/dowhy/dowhy_ihdp_data_example.html

3.2 Fuzzy Causal Effect Variational Autoencoder (FCEVAE-V2) Second Architecture

In order to create an architecture capable of dealing with noisy and missing data, we created FCEVAE-V2 architecture by integrating our fuzzy rules from [3] into the CEVAE's loss function. Our architecture is divided into two main components:

Figure 2. Part A: a conditional autoencoder that randomly generates equally unbiased samples from a dataset.

Figure 2. Part B: it takes the input from the previous step and uses fuzzy causal rules integrated into CEVAE's loss function to extract causal relationships.

We will briefly explain our architecture steps in the following:

Figure 2. Part A:
Before explaining **Part A** in detail, we briefly explain the difference between Variational Autoencoders (VAE) and the conditional variational autoencoder (CVAE) we used in Figure 2. Part A. Whereas the VAE architecture does not apply any condition during sampling from datasets, CVAE uses the conditioning method for the sampling process [21, 22].

The main goal behind the step A is to generate unbiased equal samples without missing data. To do so, Conditional VAE (Figure 2 A) takes a dataset with missing data and generates equal amount of sampling from conditional distribution of the dataset's columns. That is, we create a condition matrix (for which its columns are the output of the Conditional VAE that generates un-biased samples) so that it removes the missing

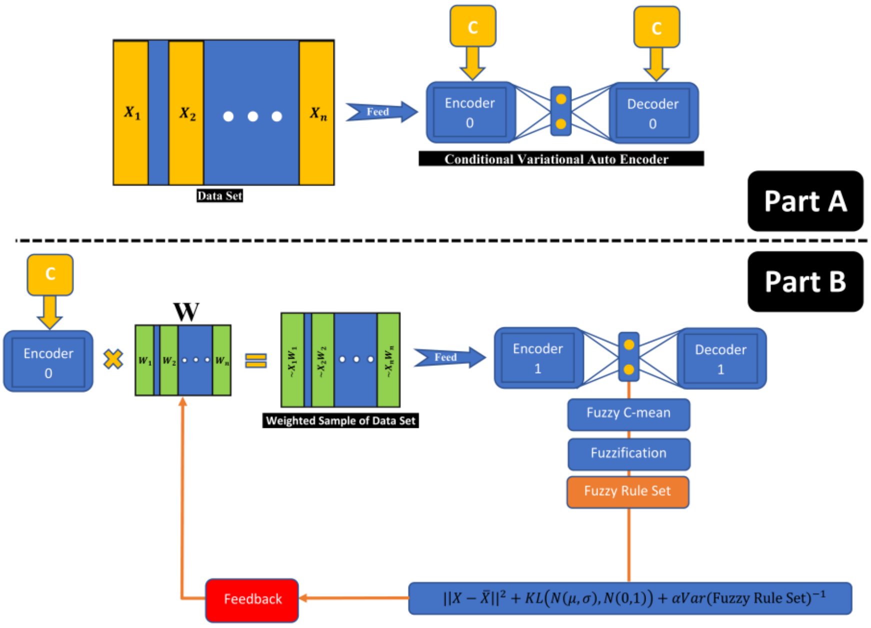

Figure 2. FCEVAE-V2 where probabilistic Fuzzy logic rules are integrated with the CEVAE loss function.

data's bias ratio. For example, assume that for a given dataset $D = (X_0, \dots, X_n)$, where X_is are the columns with length $l(X_i)$. We have $M = (m_0, \dots, m_n)$ where m_is are the corresponding missing data ratio. We generate a condition matrix $\mathbb{C} = (C_0, \dots, C_n)$, such that C_is are binary vectors with length $l(X_i)$. If the corresponding dataset's element is missing, each element of C_i, such as c_{ij}, equals 0. Otherwise the value is equal to 1.

$$\mathbf{E}[\log(\mathbf{X}\,|\mathbf{z}\,,\mathbb{C})] - \mathbf{D_{KL}}\,[\mathbf{Q}(\mathbf{z}\,|\mathbf{X}\,,\mathbb{C})\ \|\ \mathbf{P}(\mathbf{z}\,|\,\mathbb{C})] \qquad \textit{Equation 1}$$

Equation 1 is the CVAE's objective function. Q and P are the conditional distribution of CVAE's encoder and decoder respectively. KL is the KL divergence. The model learns P and Q given the condition matrix $\mathbb{C}$ [21, 22].

Figure 2. Part B: Part A's output is an unbiased sample S with no missing values. In Part B, we create a matrix W such that its columns will show possible causal relationships between the dataset D's columns (see Table 4). That is, once we calculate W, a higher value in a column (i.e., gestat10 in Table 4) shows a higher influence of the column on the outcome (see Table 4). We must emphasize that contrary to the previous works that used gestat10 as the cause, we used *all* columns to calculate possible causes.
To do so, we first initialized the randomly generated matrix W with size $(n * n)$ where n is the number of the D columns. An important note is that since the matrix W's values are randomly generated, for different executions we get slightly different values for ITE, ATE and the values in Table 4. We then multiply W (see our above SPFL theory [17]) by the output from Part A. The result of the previous step feedforward into the CEVAE (**Figure 2.B**).
After encoding the data in Figure 2, Part B's encoder section, the resulting data is partitioned using the Fuzzy-C mean [1] algorithm. This partitioning is done to automatically find fuzzy membership intervals without the need for an expert to define them. We then use fuzzy rules from [3] to fuzzify the result. The next step is to add fuzzy rules to the CEVAE loss function (Equation 2):

$$\left|\left|\mathbf{X}-\overline{\mathbf{X}}\right|\right|^{2}+\mathbf{KL}\big(\mathbf{N}(\boldsymbol{\mu},\boldsymbol{\sigma}),\mathbf{N}(\mathbf{0},\mathbf{1})\big)+\boldsymbol{\alpha}\mathbf{Var}(\mathbf{FuzzyRuleset})^{-1} \qquad \textit{\textbf{Equation 2}}$$

In the above equation, the first term is the reconstruction error. The second term is KL divergence. The third term calculates the variance of the fuzzy rule set according to the association and causal fuzzy rules from [1]. $\alpha \in [0, 1]$ is the training hyper-parameter. It helps the model include the influences of the fuzzy ruleset from [3] to the loss function. For an $\alpha = 0$, we have the original CEVAE architecture.
The above loss function's output is passed to the back-propagation algorithm to update the W (Figure 2 Part B, red rectangle).

Noise	$N \sim N(0.10, 0.5)$	$N \sim N(0.15, 0.5)$	$N \sim N(0.20, 0.5)$
$ATE_{\mathbf{FCEVAE\_V2}}$	3.27542	2.65458	1.76581
$ATE_{\mathbf{CEVAE}}$	3.35354	2.61252	1.91257
$ATE_{\mathbf{Dowhy}}$	1.99664	1.37661	1.28898

Table 2: FCEVAE-V2 performance on the noisy IHDP dataset.

The updated Ws are multiplied by the output from Part A. Again, the result are passed to the FCEVAE-V2 where the model applies C-mean and fuzzification and calculates fuzzy loss function before using back-propagation algorithm. FCEVAE-V2 continues the above steps until the result converges to a minimum value for the loss function.

3.3 Second Architecture's Experiments

Similar to the CEVAE projet [2], and the DoWhy project [19], we tried FCEVAE-V2 with the IHDP [9] and TWINS [2] datasets. With the TWINS dataset, the goal is to find the possible causal relationships between the weight of twins and their death rate. The main difference between FCEVAE-V2, CEVAE, and DoWhy architectures is that while other architectures add noise to one specific column (gestat10 column), we added noises to the whole dataset. We did this to show that DLs equipped with non-classical logic rules are tolerant to multiple noise source.

After applying FCEVAE-V2 to the IHDP dataset, we obtained similar ATE and ITE values to the CEVAE and Dowhy's project's outputs (see GitHub). To try FCEVAE-V2 with noisy data, similar to [2], we applied the gaussian noise $N \sim N(\mu, 0.5)$ where $\mu \in (0.10, 0.15, 0.20)$ on the IHDP dataset and passed it to FCEVAE-V2 in order to measure the network's noise tolerance level. Table 2 shows that compared to other architectures, our architecture is more tolerant to noises (a lower ATE is better).

We also applied the noise to the TWINS dataset and passed the noisy data to FCEVAE-V2. Table 3 shows that comparing to CEVAE and DoWhy, our model gives lower ATE values.

Noise	$\boldsymbol{N \sim N(0.10, 0.5)}$	$\boldsymbol{N \sim N(0.15, 0.5)}$	$\boldsymbol{N \sim N(0.20, 0.5)}$
$\boldsymbol{ATE_{\mathrm{FCEVAE\_V2}}}$	-0.02616	-0.02711	-0.05121
$\boldsymbol{ATE_{\mathrm{Dowhy}}}$	-0.06901	-0.11760	-0.17351
$\boldsymbol{ATE_{\mathrm{CEVAE}}}$	-0.02720	-0.02931	-0.06245

Table 3: FCEVAE_V2 performance on noisy TWINS dataset.

Similarly to the CEVAE [2] and DoWhy projects, we used the TWINS dataset with FCEVAE-V2. It must be noted that in the previous works the authors only used gestat10 column to calculate the possible cause of the twins' death rate. In this study we used all columns with FCEVAE-V2. Table 4 shows the most important relationships between columns (to see the full result, the reader is referred to [5]). We would like to remind the reader that although we used a heatmap to show the values in Table 4, these values are not correlations and/or covariance matrices. These values are the final values of the matrix W (see above), and they were obtained after using CEVAE's probability approach and many iterations of the c-mean clustering algorithm, fuzzification, and fuzzy rule integration to the CEVAE cost function.

In table 4, all values belong to the [0 ,1] interval. The higher value shows the stronger possible cause between columns. For instance, similar to [2], our model revealed a

[5] To see the complete result please see https://github.com/joseffaghihi/Causal-fuzzy-CEVAE/tree/main/2021-12-14

strong relationship (0.52%, please see the outcome row in Table 4) between GESTAT10 and outcome which is one of the highest values in the outcome row. That is, the GESTAT10 column influences many other columns in the dataset (see the high values in the GESTAT10 column).

Table 4. FCEVAE-V2 output for TWINS dataset. Each element $\epsilon[0,1]$ interval is the causality level of the associated columns and rows from matrix W. The dark blue color shows possible causal relationships. Values can be read by zooming into the table (high resolution table)

Limitations: Similar to previous work, FCEVAE-V2 is capable of finding the causal relationships between the TWINS dataset columns (TWINS' description). Since our model uses all columns in TWINS, it also found other possible causal relationships between columns that were not mentioned in previous works.

We have found some health-related papers that could potentially suggest a scientific foundation for the results generated by our model. For instance:

Column name and description		
mpre5 (trimester prenatal care begun)	adequacy (adequacy of care)	[23]
mpre5 (trimester prenatal care begun)	Eclamp (risk factor, Eclampsia)	[24]
mpre5 (trimester prenatal care begun)	Incervix (risk factor, Incompetent cervix)	[25]

Table 5. TWINS data set columns and their description according to TWINS' description

However, this is only the very surface of what needs to be done next. Given that FCEVAE-V2 uses both probability and fuzzy approaches to calculate the casual relationship between columns in the dataset, at this point, we cannot provide an explanation for how these values are calculated precisely. We aim to do so in our future work. We also encourage the readers to contact us, should they find any explanation for our result (the code is on GitHub [5]).

4 Conclusion

In this paper, we have shown that Deep Learning algorithms (DLs) equipped with non-classical logics such as PFLs are capable of reasoning with multiple sources of missing data or noise. This had not been done in previous works; only one source of noise was previously used.
To do so, we created two architectures: 1) First, after applying probabilistic fuzzy logic association and causal rules (PFLs) to the dataset, the architecture feedforwarded the output to the Causal Variational Autoencoders (CEVAE) architecture [2]; 2) Second, we integrated PFLs into the CEVAE's loss function. Compared to the Microsoft DoWhy, and the original CEVAE architecture, our FCEVAE-V2 is more tolerant to datasets with missing data and multiple sources of noise.
In contrast to the original CEVAE architecture which relies heavily on the treatment column to be determined by human experts, our FCEVAE-V2 does it automatically. That is, in order to reveal possible causal relationships between columns, our model applies causal rules to all columns. To prevent combinatorial problems when selecting treatment FCEVAE-V2 uses the CEVAE compression technique.
Much work remains to be done. An important limitation of our work is explaining the calculations of the causal relationships between columns, and their interpretation in real-life scientific contexts.

Using the FCEVAE-V2, reveal the brain's different parts communication causal pattern.
The output of the FCEVAE-V2 (please see the code). Shows, the brain uses 68, 72 ,75 ,81 and 95 hz frequencies to establish a connection between far lobes (frontal and occipital)

Acknowledgements:

We thank Sioui Maldonado Bouchard for kindly proofreading this paper.

References

1. Bezdek, J.C., Ehrlich, R., and Full, W., *FCM: The fuzzy c-means clustering algorithm.* Computers & geosciences, 1984. **10**(2-3): p. 191-203.
2. Louizos, C., Shalit, U., Mooij, J.M., Sontag, D., Zemel, R., and Welling, M., *Causal effect inference with deep latent-variable models*, in *Advances in Neural Information Processing Systems*. 2017. p. 6446-6456.
3. Faghihi, U., Robert, S., Poirier, P., and Barkaoui, Y., *From Association to Reasoning, an Alternative to Pearl's Causal Reasoning.* In Proceedings of AAAI-FLAIRS 2020. North-Miami-Beach (Florida), 2020.
4. Faghihi, U., Maldonado-Bouchard, S., and Biskri, I., *Science of data: from correlation to causation.* Springer, 2021.

5. An, J. and Cho, S., *Variational autoencoder based anomaly detection using reconstruction probability.* Special Lecture on IE, 2015. **2**(1): p. 1-18.
6. Yao, L., Li, S., Li, Y., Huai, M., Gao, J., and Zhang, A., *Representation learning for treatment effect estimation from observational data.* Advances in Neural Information Processing Systems, 2018. **31**.
7. Guo, R., Cheng, L., Li, J., Hahn, P.R., and Liu, H., *A survey of learning causality with data: Problems and methods.* ACM Computing Surveys (CSUR), 2020. **53**(4): p. 1-37.
8. Roeder, G., Metz, L., and Kingma, D. *On linear identifiability of learned representations*. in *International Conference on Machine Learning*. 2021. PMLR.
9. Hill, J.L., *Bayesian nonparametric modeling for causal inference.* Journal of Computational and Graphical Statistics, 2011. **20**(1): p. 217-240.
10. Chipman, H.A., George, E.I., and McCulloch, R.E., *BART: Bayesian additive regression trees.* The Annals of Applied Statistics, 2010. **4**(1): p. 266-298.
11. Shalit, U., Johansson, F.D., and Sontag, D. *Estimating individual treatment effect: generalization bounds and algorithms*. in *International Conference on Machine Learning*. 2017. PMLR.
12. Khemakhem, I., Kingma, D., Monti, R., and Hyvarinen, A., *Variational autoencoders and nonlinear ica: A unifying framework*, in *International Conference on Artificial Intelligence and Statistics*. 2020, PMLR. p. 2207-2217.
13. Wu, P. and Fukumizu, K., *Intact-VAE: Estimating Treatment Effects under Unobserved Confounding.* arXiv preprint arXiv:2101.06662, 2021.
14. Yager, R.R. and Zadeh, L.A., *An introduction to fuzzy logic applications in intelligent systems*. Vol. 165. 2012: Springer Science & Business Media.
15. Zhao, D.-M., Wang, J.-H., Wu, J., and Ma, J.-F. *Using fuzzy logic and entropy theory to risk assessment of the information security*. in *2005 International Conference on Machine Learning and Cybernetics*. 2005. IEEE.
16. Cheng, P.-C., Rohatgi, P., Keser, C., Karger, P.A., Wagner, G.M., and Reninger, A.S. *Fuzzy multi-level security: An experiment on quantified risk-adaptive access control*. in *2007 IEEE Symposium on Security and Privacy (SP'07)*. 2007. IEEE.
17. Saki, A. and Faghihi, U., *Fuzzy Rule Based Probability Theory (IN PREPARATION)* 2022.
18. Ng, A., O'Reilly, and Associates, *AI is the New Electricity*. 2018: O'Reilly Media.
19. Sharma, A. and Kiciman, E., *DoWhy: An end-to-end library for causal inference.* arXiv preprint arXiv:2011.04216, 2020.
20. Robert, S., Faghihi, U., Barkaoui, Y., and Ghazzali, N. *Causality in Probabilistic Fuzzy Logic and Alternative Causes as Fuzzy Duals*. in *International Conference on Computational Collective Intelligence*. 2020. Springer.
21. Sohn, K., Lee, H., and Yan, X., *Learning structured output representation using deep conditional generative models.* Advances in neural information processing systems, 2015. **28**: p. 3483-3491.
22. Doersch, C., *Tutorial on variational autoencoders.* arXiv preprint arXiv:1606.05908, 2016.
23. Tayebi, T., Zahrani, S.T., and Mohammadpour, R., *Relationship between adequacy of prenatal care utilization index and pregnancy outcomes.* Iranian journal of nursing and midwifery research, 2013. **18**(5): p. 360.
24. Herrera, J., Chaudhuri, G., and López-Jaramillo, P., *Is infection a major risk factor for preeclampsia?* Medical hypotheses, 2001. **57**(3): p. 393-397.
25. Nicolaides, K.H., *Turning the pyramid of prenatal care.* Fetal diagnosis and therapy, 2011. **29**(3): p. 183-196.