Deep Learning of Markov Model Based Machines for Determination of Better Treatment Option Decisions for Infertile Women

Arni S.R. Srinivasa Rao, PhD^{a*, b, c,}
Michael P. Diamond, MD^{d, e}

Abstract. *In this technical article, we are proposing ideas those we have been developing of how machine learning and deep learning techniques can potentially assist obstetricians / gynecologists in better clinical decision making using infertile women in their treatment options in combination with mathematical modeling in pregnant women as examples.*

Key words: Deep learning, state-spaces, AI in medicine

MSC: 68T99, 60J10, 60J22

Machine learning approaches in medicine, where applicable, will have tremendous impact on monitoring patients' health, especially, when care providers want to do the following: improve the precision of the treatment impacts, reduce human errors that could be due to repeated and

routine human actions, provide consistency in treatment prescriptions, or any combinations of these. In this commentary, we highlight machine learning (ML) approaches and Artificial Intelligence (AI) based decisions that we believe will have increasing prevalence in clinical care throughout our discipline and others. For the purpose of this discussion, we provide example of infertility treatment for assisting in patients' healthcare when they present to them obstetrician/gynecologist.

Infertility Care Options

Suppose we develop an algorithm that can be used in arriving at a decision on which of the treatment options are appropriate for an infertile couple, that can maximize her probability of delivering a child. The data that would be fed into this algorithm usually consists of clinical, demographic and genetic level information from an individual infertile couple for whom a decision on treatment options is to be provided by algorithm. This information we call patients' level data.

Machines using these algorithms are equipped with parameters whose values were computed with prior information on large population level historical infertile couples' data who successfully delivered a child with one of the treatment options available, say either IVF or through ovarian stimulation with ovulation induction medicines. There are also types of ovulation induction medicine whose details are fed into the algorithm to be used in the decision making process. These parameters are associated with probabilities of conception, and then for delivery of a baby for various types of infertile couple. One of the methods to obtain such parameters is through Markov Chain based transition probabilities matrices (1-3). Applications of Markov modeling for infertility related data can be found in (4-5) and historically original applications of stochastic processes in biological sciences can be read in (6-8). See Appendix II for the proposed method of computing parameters through Markov Chains approach.

^a Laboratory for Theory and Mathematical Modeling, Division of Infectious Diseases, Medical College of Georgia, Augusta University, USA. Address: 1120, 15th Street, Augusta, GA, 30912. *(Corresponding) Email: arrao@augusta.edu

^b Division of Epidemiology-Department of Population Health, Medical College of Georgia, Augusta University, USA

^c Department of Mathematics, College of Science and Mathematics, Augusta University, USA

^d Department of Obstetrics and Gynecology, Medical College of Georgia, Augusta University, USA. Address: 1120, 15th Street, CJ-1306, Augusta, GA, 30912. Email: midiamond@augusta.edu

^e Senior Vice President, Office of Senior Vice President, Augusta University, USA

Figure 1. Artificial Intelligence is manufactured from deep learning algorithms or machine learning algorithms or from the both

Machines which have the above mentioned algorithms with parameters values obtained from treatment outcomes of a large number of infertile couple we call machine learning algorithm for fertility treatment outcome (MLAFTO). MLAFTO will predict the chances of conceiving and delivering for a new couple who have come to an infertility clinic. MLAFTO would be able to compute probabilities for all the combinations of patient level characteristics combined with their treatment options (See Table 1 for an idea about combination of patient data with treatment options).

Once the chances of predicting the outcomes are computed (see the Appendix for details), then through induction algorithm (i.e. from known values of probabilities of outcomes) the lowest probability of delivering a baby corresponding to a specific treatment option combination, to the highest such probability among all possible sets of probabilities, are sorted out. This procedure through the use of an induction algorithm to suggest a best treatment option we call here as AI based treatment option for an infertile couple.

Table 1 is inserted here (see end of the document)

The Artificial Intelligence quotient of a machine assists in decision making processes for deciding which treatment option(s) are best suited for an infertile woman or a couple who comes to an infertility clinic (based on large

historical data and the data on new couple). There are design issues involved on the type of information for the past data which we want to use. That is, whether to use deep learning of the historical data before AI based decision is implemented, or to use machine learning algorithm developed on the past data before AI based decision making can be initiated. AI can be manufactured from deep learning or from machine learning algorithms (Figure 1). In deep learning, each time an infertile couple comes to a clinic, AI will get freshly manufactured from all the data that was fed into the MLAFTO prior to this woman (or couple) and then assists to arrive at a conclusion on which of the treatment options are best suited for the couple. See Appendix III for the key differences between two approaches. That is, the probabilities of conceiving and delivering a baby are computed afresh based on all the infertile couple treatment data (as in Table 1) prior to the new couple who came for infertility care. When we have decided to go with MLAFTO (without deep learning), a fixed set of past data on the infertile woman (or a couple) (as shown in Table 1) will be used to obtain probabilities of conception and delivering a live birth, and then a newly arrived couple's characteristics will be matched with existing combinations as in Table 1 to suggest a best treatment option for the couple. That is, in the machine learning approach, probabilities of conceiving and for a live birth are computed *a priori* (from a fixed historical data without any further updates), and when a new couple comes for an infertility care, then new couple's characteristics are matched with existing characteristics of the fixed historical data for which *a priori* probabilities were computed. Through a step by step matching of new couple characteristics with the existing characteristics (inductive reasoning) the closest characteristics that match to the past data is decided. For this closest match the corresponding probabilities computed *a priori* are suggested as possible probabilities for the new couple.

Figure 2. Disjoint sets of infertile women across all the treatment options within the background characteristics (B_*).

Examples of clinical variables include, the duration of infertility, information on earlier treatments if any, depression, anxiety, etc, Potential demographic variables are, namely, age, level of education, duration of relation between couple who came for treatment, etc, genetic variables include: family history of infertility, family history of abortion, miscarriage, etc, other variables include, financial status of the couple, income level of the couple, etc,

Deep learning updates past treatment effects data continuously whereas machine learning computes parameters based on fixed data. However, machine learning parameters can be also updated as newer data becomes available (and also should there be a need to update parameters those were already computed due to a change in treatment options or for some other reason).

Various information on clinical variables and respective treatment outcomes of the past indicated across the big data sets will need to be continuously updated to construct intelligence on the future treatment options. This continuous updating is done wither through pre-constructed models or through deep learning. Some of the several variables like anxiety, depression, level of education, genetic information could vary across the populations in the world, and simplification of their role in treatment effects could lower the chances of creating precise intelligence and for predicting the capacity of treatments of various

populations across the world. We suggest that the information on the variables needed to be shared across various populations for capturing better intelligence. The flexibility of the models can be incorporated through model selection procedures to avoid bias. The advantages of deep learning will be maximized with the usage of big data sets and the conception and baby delivery probabilities can be precisely predicted. A word of caution is that deep learning on big data would not automatically guarantee to gather precise AI for predicting the treatment effects and these big data have to be structured accurately with mathematical constructions to be able to build models.

Acknowledgements: We thank the following individuals in the alphabetical order of their last name for very valuable comments: Medina Jackson-Browne (Brown University, Providence), N.V. Joshi (Indian Institute of Science, Bangalore), K. Praveen (Microsoft, Irvine), P. Sashank (CEO, Exactco, Hyderabad). Actual journal submission contain more examples and results.

Appendix I: Machine Learning Algorithm for Fertility Treatment Outcome (MLAFTO)

Let $x(\{c_*, d_*, g_*, o_*\})$ be the probability that an infertile woman with characteristics $*$ such that $* = 1, 2, \dots, k \times l \times m \times n$ with i^{th} -ovulation induction (OI) treatment or IVF for $i = 1, 2, \dots, p$ is conceived or delivered.

We compute x values for all $*$ combinations.

We compute $\max_i(\max_* x)$, $\max_*(\max_i x)$. Once the maximum probabilities are computed and ranking of these probabilities over various $*$ and i will provide relative chances of conceiving and delivering a live birth. AI quotient will match these combinations of $*$ and i is with the new couple who come to the clinic and suggest the probability of conceiving and delivering a live birth. See Appendix II for computing probabilities and descriptions related to $\max$ functions.

Appendix II: Computation of Probabilities Through Markov Chains

In this Appendix, we propose a Markov Chain based approach in computing

probabilities of conception and delivering a live birth under various treatment options.

Suppose we want to compute the probability of conception and then delivering a baby for an infertile woman with a combination of background variables, say, $\{c_5, d_4, g_1, o_1\}$ and with OL_i or IVF treatments explained in the paper. Let B_1 be the set of all infertile women with background variables $B_1 = \{\{c_5, d_4, g_1, o_1\}\}$ who will be on OL_i or IVF treatment options. Let $x(B_1)_{jc}^{(T)}$ be the probability that an infertile woman at the state j with characteristics $\{c_5, d_4, g_1, o_1\}$ and with i^{th} -ovulation induction (OI) treatment or IVF for $i = 1, 2, \dots, p$ is conceived in T - time steps. Let $x_i(B_1)_{jb}^{(T)}$ be such a probability to deliver a baby, and $x(B_1)_{cb}$ be the probability of baby is born to a woman within B_1 given that the woman is conceived. These three probabilities can be computed using below formulas:

$$x_i(B_1)_{jc}^{(T)} = \frac{\int_{s \in B_1} W_{i,T}^{j \rightarrow c} ds}{\int_{s \in B_1} W_i^j ds} \dots (A_{2.1})$$

$$x_i(B_1)_{jb}^{(T)} = \frac{\int_{s \in B_1} W_{i,T}^{j \rightarrow b} ds}{\int_{s \in B_1} W_i^j ds} \dots (A_{2.2})$$

$$x_i(B_1)_{cb} = \frac{\int_{s \in B_1} W_{i,T}^{c \rightarrow b} ds}{\int_{s \in B_1} W_i^c ds} \dots (A_{2.3})$$

where $W_{i,T}^{j \rightarrow c}(s)$ denotes s^{th} infertile woman in the state j who is on i^{th} treatment conceives in T - time steps, $W_i^j(s)$ denotes s^{th} infertile woman in the state j who is on i^{th} treatment. $\int_{s \in B_1} W_{i,T}^{j \rightarrow c}(s) ds$ is the total number of infertile women in the set B_1 who have moved from the state j to the state c who are on i^{th} treatment and $\int_{s \in B_1} W_i^j(s) ds$ is the total number of women in the set B_1 who are at the state j who are on i^{th} treatment. Suppose B_2 be another set of all infertile women with different background variables, say, $B_2 = \{\{c_1, d_3, g_6, o_2\}\}$ with OL_i or IVF treatments, then we can compute

corresponding transition probabilities by a similar type of formulas as in $(A_{2.1}) - (A_{2.2})$.

Probability of transition from the state c to the state b does not depend upon $\int_{s \in B_1} W_i^j(s) ds$ but only on the $\int_{s \in B_1} W_i^c(s) ds$, so the random variable responsible for the transition between these two states, say Y , obeys Markov property. Moreover, the transition probability matrix $P_i(B_1)$ for the set of infertile women B_1 between states $\{j, c, b\}$ who are on i^{th} treatment can be written as,

$$P_i(B_1) = \begin{bmatrix} & j & c & b \\ j & x_i(B_1)_{jj} & x_i(B_1)_{jc} & x_i(B_1)_{jb} \\ c & x_i(B_1)_{cj} & x_i(B_1)_{cc} & x_i(B_1)_{cb} \\ b & x_i(B_1)_{bj} & x_i(B_1)_{bc} & x_i(B_1)_{bb} \end{bmatrix}$$

where $x_i(B_1)_{jj} + x_i(B_1)_{jc} = 1$, $x_i(B_1)_{cc} + x_i(B_1)_{cb} = 1$ and $x_i(B_1)_{bb} = 1$. $x_i(B_1)_{jb} = 0$ due to Markov property. Where as $x_i(B_1)_{cj} = x_i(B_1)_{bj} = x_i(B_1)_{bc} = 0$ due to transition from $c \rightarrow j$, $b \rightarrow j$ and $b \rightarrow c$ are impossible. Similarly, we will compute

$$P_i(B_*) = \begin{bmatrix} & j & c & b \\ j & x_i(B_*)_{jj} & x_i(B_*)_{jc} & x_i(B_*)_{jb} \\ c & x_i(B_*)_{cj} & x_i(B_*)_{cc} & x_i(B_*)_{cb} \\ b & x_i(B_*)_{bj} & x_i(B_*)_{bc} & x_i(B_*)_{bb} \end{bmatrix}$$

for $* = 1, 2, \dots, k \times l \times m \times n$. Let $W_i^j(B_*)$ be the number of infertile women (state j) within background characteristics B_* who are on i^{th} treatment and let $W^j(B_*)$ be the total number of infertile women within background characteristics B_* such that

$$W^j(B_*) = \bigcup_{i=1}^p W_i^j(B_*)$$

$$\bigcap_{i=1}^p W_i^j(B_*) = \emptyset \text{ (empty set).}$$

Once $P_i(B_*)$ is computed based on certain design of the sample population, the sizes of $W_i^j(B_*)$ are not changed for computing probabilities using $(A_{2.1}) - (A_{2.2})$. That is, the

matrix $P_i(B_*)$ is not updated based on newer women who have started treatment after the designed time interval.

Two functions: $\max_i \{x_i(B_*)_{jc}\}$ and $\max_i \{x_i(B_*)_{cb}\}$

The function $\max_i \{x_i(B_*)_{jc}\}$ describes that the maximum of the probability values of women with background characteristics B_* across all the treatments, which is obtained as

$$\max \left\{ \frac{\int_{s \in B_*} W_{1,T}^{j \rightarrow c}(s) ds}{\int_{s \in B_*} W_1^j(s) ds}, \frac{\int_{s \in B_*} W_{2,T}^{j \rightarrow c}(s) ds}{\int_{s \in B_*} W_2^j(s) ds}, \dots \right\} \dots (A_{2.4})$$

Through the expression $(A_{2.4})$ we will obtain $k \times l \times m \times n$ maximum values, where each maximum value represents maximum probability of conceiving by an infertile woman from a particular set of background characteristics and corresponding treatment type for which this maximum value is obtained. Similarly, we can construct $\max_i \{x_i(B_*)_{cb}\}$. The function $\max_i \{x_i(B_*)_{cb}\}$ describes that the maximum probability of conceiving within the women who are i^{th} treatment across different background characteristics, which is obtained as

$$\max \left\{ \frac{\int_{s \in B_1} W_{i,T}^{j \rightarrow c}(s) ds}{\int_{s \in B_1} W_i^j(s) ds}, \frac{\int_{s \in B_2} W_{i,T}^{j \rightarrow c}(s) ds}{\int_{s \in B_2} W_i^j(s) ds}, \dots \right\} \dots (A_{2.5})$$

Let $W_i^j(B_*)$ be the number of infertile women (state j) within background characteristics B_* who are on i^{th} treatment and let $W^j(B_*)$ be the total number of infertile women within background characteristics B_* , then

$$W^j(B_*) = \bigcup_{i=1}^p W_i^j(B_*)$$

$$\bigcap_{i=1}^p W_i^j(B_*) = \emptyset \text{ (empty set).}$$

See also Figure 2 to see this disjoint property of infertile women within each background characteristics.

Result: Total infertile women with background characteristics $\{B_*\}$ can be written as the union of disjoint sets of women across all treatment options, i.e.

$$\int_{s \in B_*} W_i^j(s) ds = \bigcup_{i=1}^p W_i^j(B_*) ds \dots \dots (A_{2.4})$$

Appendix III: Machine Learning Versus Deep Learning in Computing Probabilities of Conception and Delivery

Suppose a new infertile woman whose background characteristics $\{B_N\}$ is interested to start one of the available treatments OL_i or IVF . Let us understand how machine learning techniques are applied to decide which of the treatment will give maximum chance of conception and delivering a baby. Prior to a decision making process on treatment options for this woman, let us suppose that probabilities of conception and delivery were previously computed through MLAFTO explained in the Appendix I and $P(B_*)$ for all $*$ in the Appendix II. The data used for these two computations is usually a predetermined or pre-designed one, i.e. the time frame and other design aspects of the data were well defined and are without any data related errors. MLAFTO matches the new infertile woman characteristic set B_N with the sets $\{B_* : * = 1, 2, \dots, k \times l \times m \times n\}$. Let $\{B_y\}$ be the set that matches with the new woman characteristics such that $\{B_y\} - \{B_N\} = \emptyset$ (null set). The corresponding values of

$$x(B_y)_{jc} \text{ and } x(B_y)_{cb}$$

are considered as chances of conception and chances of delivery for the new woman who came to the clinic.

Note that, the success or failure data of woman with $\{B_N\}$ is not used in computation of $P(B_*)$ for all $*$ which is the key for machine learning type of algorithm.

If each treatment trial of a woman whether or not that woman conceives is considered as one time step of treatment (or 1-

cycle of treatment) and the duration from conceiving of a woman to whether or not a baby is delivered is considered as one time step of pregnancy (or 1-cycle of pregnancy), and let $x(B_y)_{jc}^{(n)}$ and $x(B_y)_{cb}^{(n)}$ be the corresponding n – step or n – cycle probabilities, then by Markov property, we have

$$x(B_y)_{jc}^{(n)} \times x(B_y)_{cb}^{(m)} = x(B_y)_{jb}^{(n+m)}$$

When another infertile woman with background characteristics $\{B_M\}$ comes to the clinic for the purpose of decision making of which type of treatment will be needed for a successful delivery, the prior computed transition probability matrix $P(B_*)$ for all $*$ that was used in matching for a woman with $\{B_N\}$ was not updated with the success or failure information of the woman with $\{B_N\}$. In a way, the matrices $P_i(B_*)$ are static in case we are using machine learning algorithms and these are not influenced by new data generated on newer infertile women who come to the clinic after constructing $P_i(B_*)$.

Once an infertile woman walks into the clinic with background characteristics $\{B_M\}$, if deep learning techniques are implemented to predict the probabilities of conceiving (say, $y(B_N)_{jc}$) and the delivery (say, $y(B_N)_{cb}$), then the computations of such probabilities are different than machine learning techniques. Each time a new infertile woman with $\{B_M\}$ comes to the clinic for the treatment purposes, instead of matching procedure with the existing static model explained above, deep learning involves reconstructing of the transition probability matrices $P_i(B_*)$, for $i = 1, 2, \dots, p$ for conceiving and delivery with whatever data that is available prior to arriving of the woman with $\{B_M\}$. Rest of the computational procedures explained in the Appendix II remains the same. Deep learning techniques usually delay the output due to reconstructing of the $P_i(B_*)$ each and every time a new infertile woman comes to clinic.

General introductions of machine learning techniques, motivations, and key ideologies were explained in a variety of research areas can be found in (9-13). Specific

ideas related to deep learning techniques were also well developed (14), deep learning techniques and applications were summarized (15) and an overview of importance of machine learning algorithms in medicine can be found in (16). As explained in our article, the machine learning and deep learning techniques broadly use the same data within the specific goals, but their approach of handling the data and models distinguish them from each other. Statistical thinking had contributed several aspects of machine learning, for example, in developing computationally intense data classification algorithms, methods in data search and matching probabilities, data mining techniques, model classification and model fitting algorithms, and a combination of all these (see for example, (17-29) and for a collection of articles related to statistical methods in machine learning see (30). Model-based machine learning methods (31) and the construction of coefficients in a regression model can be benefited by machine learning methods (32).

Deep learning techniques, instead of focusing on model-based approaches, they would assist in understanding intricate structures of the large data sets and various interlinkages between these data sets (33). Importance of unsupervised pre-training to the structural architecture and the hypothesis of testing design effects of such experiments are well studied (34, 35). Deep learning and machine learning techniques could also assist in questions related to health informatics, disease detection, item response theories, and bioinformatics research (36-41). There were also successful methods in deep learning algorithms which score patients in ICU (Intensive Care Unit) for their severity and to predict mortality without using any model based assumptions in scoring systems (42) and for other medical applications, for example detection of worms through endoscopy (43), ophthalmology studies (44), cardiovascular studies (45), Parkinson's disease data (46), medical scoring systems (47). Deep learning procedures involved in various levels of abstraction for ranking system models can be found in (48, 49), applications for mathematical models, parameter computations and stability of algorithms are found in (50-56).

Statistical and stochastic modeling principles were applied in deep learning algorithms to strengthen the object search capabilities or for improved model fitting in uncertainty (57, 59). Boltzmann machines assist in the deep understanding of the data by linking layer level structured data and then by estimating model parameters through maximum likelihood methods (60, 61). Random backpropagation and backpropagation methods help in stochastics transition matrix formations and computing quicker search algorithms in higher dimensional stochastic matrices and literature related to backpropagation could be found in several places, for example, see in (62-65). A survey of statistical learning algorithms and their performance evaluations can be found in (66).

Appendix V: Theorems

Theorem A.1: When W^j is the total number of infertile women (state j) whose data is used in the machine learning algorithm and $\delta \in [1, klmn]$ and $\alpha \in [1, p]$ are considered as continuous for background characteristics and treatment options, then

$$\frac{1}{pklmn} \left[\int_{\delta=1}^{klmn} \int_{\alpha=1}^p \frac{W_{\alpha}^{j \rightarrow c}(B_{\delta})}{W_{\alpha}^j(B_{\delta})} d\alpha d\delta + \int_{\delta=1}^{klmn} \int_{\alpha=1}^p \frac{W_{\alpha}^{c \rightarrow b}(B_{\delta})}{W_{\alpha}^j(B_{\delta})} d\alpha d\delta \right] \leq 1$$

Proof: We have,

$$W^j(B_{\delta}) = \int_1^p \frac{W_{\alpha}^j(B_{\delta})}{W^j(B_{\delta})} d\alpha \dots \dots (A_{5.1})$$

and

$$W^j = \int_1^{klmn} \int_1^p \frac{W_{\alpha}^j(B_{\delta})}{W^j(B_{\delta})} d\alpha d\delta \dots \dots (A_{5.2})$$

Note that,

$$\frac{W_1^{j \rightarrow c}(B_{\delta})}{W_1^j(B_{\delta})} + \frac{W_2^{j \rightarrow c}(B_{\delta})}{W_2^j(B_{\delta})} + \dots + \frac{W_p^{j \rightarrow c}(B_{\delta})}{W_p^j(B_{\delta})} \leq p \dots \dots (A_{5.3})$$

and

$$\frac{W_1^{c \rightarrow b}(B_{\delta})}{W_1^j(B_{\delta})} + \frac{W_2^{c \rightarrow b}(B_{\delta})}{W_2^j(B_{\delta})} + \dots + \frac{W_p^{c \rightarrow b}(B_{\delta})}{W_p^j(B_{\delta})} \leq p \dots \dots (A_{5.4})$$

from the inequality (A_{4.3}) we can obtain,

$$\int_1^{klmn} \int_1^p \frac{W_{\alpha}^{j \rightarrow c}(B_{\delta})}{W_{\alpha}^j(B_{\delta})} d\alpha d\delta \leq pklmn \dots \dots (A_{5.5})$$

from the inequality (A_{4.4}) we can obtain,

$$\int_1^{klmn} \int_1^p \frac{W_{\alpha}^{c \rightarrow b}(B_{\delta})}{W_{\alpha}^j(B_{\delta})} d\alpha d\delta \leq pklmn \dots \dots (A_{5.6})$$

Required result is deduced from two inequalities (A_{4.5}) and (A_{4.6}).

Theorem A.2: For continuous α and δ , we have

$$\left(\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta \right) \left(\int_{\delta} W_{\alpha}^b(B_{\delta}) d\delta \right) \leq \left(\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta \right)^2$$

Proof: We know,

$$\frac{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta}{\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta} = \begin{cases} 0 & \text{if no infertile women with } \alpha \text{ conceives} \\ 1 & \text{if every women with } \alpha \text{ conceives} \\ \theta \text{ for } \theta \in (0,1) & \text{if at least one woman conceives} \end{cases}$$

$$\frac{\int_{\delta} W_{\alpha}^b(B_{\delta}) d\delta}{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta} = \begin{cases} 0 & \text{if no conceived women with } \alpha \text{ delivers} \\ 1 & \text{if every women with } \alpha \text{ delivers} \\ \gamma \text{ for } \gamma \in (0,1) & \text{if at least one woman delivers} \end{cases}$$

These imply,

$$0 \leq \frac{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta}{\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta} \leq 1 \dots \dots (A_{5.7})$$

$$0 \leq \frac{\int_{\delta} W_{\alpha}^b(B_{\delta}) d\delta}{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta} \leq 1 \dots \dots (A_{5.8})$$

From $(A_{5.7})$ and $(A_{5.8})$, we can deduce required result.

Theorem A.3: Let $f: A \rightarrow \mathbb{R}^+$ and $g: B \rightarrow \mathbb{R}^+$ where A is the set of fractions of $(A_{5.7})$ and B is the set of all fractions of $(A_{5.8})$, then f and g are defined only at the adherent points of A and B , respectively.

Proof: Note that,

$$\min \left(\int_{\square} W_{\alpha}^c(B_{\delta}) d\delta \right) = \min \left(\int_{\delta} W_{\alpha}^b(B_{\delta}) d\delta \right) = 0$$

and

$$\max \left(\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta \right) = \left(\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta \right)$$

Two sets A and B are constructed from $(A_{5.7})$ and $(A_{5.8})$ as

$$A = \left\{ 0, \frac{1}{\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta}, \frac{2}{\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta}, \dots, 1 \right\} \dots \dots (A_{5.9})$$

$$B = \left\{ 0, \frac{1}{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta}, \frac{2}{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta}, \dots, 1 \right\} \dots \dots (A_{5.10})$$

From the elements of the set A as in $(A_{5.9})$, f is not defined at open subintervals

$$\left(0, \frac{1}{\int_{\square} W_{\alpha}^j(B_{\delta}) d\delta} \right), \left(\frac{1}{\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta}, \frac{2}{\int_{\delta} W_{\alpha}^j(B_{\delta}) d\delta} \right), \dots$$

and from the elements of the set B as in $(A_{5.10})$, g is not defined at open subintervals

$$\left(0, \frac{1}{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta} \right), \left(\frac{1}{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta}, \frac{2}{\int_{\delta} W_{\alpha}^c(B_{\delta}) d\delta} \right), \dots$$

Hence, f and g are defined only at the adherent points of A and B .

References:

1. McDonnell J, Goverde AJ, Rutten FF, Vermeiden JP. Multivariate Markov chain analysis of the probability of pregnancy in infertile couples undergoing assisted reproduction. *Hum Reprod.* 2002; 17(1):103-106.
2. Fiddlers AA, Dirksen CD, Dumoulin JC, et al. Cost-effectiveness of seven IVF strategies: results of a Markov decision-analytic model. *Hum Reprod.* 2009;24(7):1648-1655.
3. Rao ASRS, and Diamond MP. Role of Markov Modeling Approaches to Understand the Impact of Infertility Treatments. *Reprod Sci.* 2017 (11):1538-1543.
4. Hsieh MH, Meng MV, Turek PJ. Markov modeling of vasectomy reversal and ART for infertility: how do obstructive interval and female partner age influence cost effectiveness? *Fertil Steril.* 2007; 88(4):840-846.
5. Olive DL, Pritts EA. Markov modeling: questionable data in, questionable data out. *Fertil Steril.* 2008;89(3):746-747.
6. Bartlett, M. S. Some evolutionary stochastic processes. *J. Roy. Statist. Soc. Ser. B.* 11, (1949). 211–229.
7. Kimura M. Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci U S A.* 1955;41(3):144-150.
8. Kimura M. Some problems of stochastic processes in genetics. *Ann Math Statist.* 1957;28(4):882-901.
9. Lantz, B (2015). *Machine Learning with R: Expert techniques for predictive modeling to solve all your data analysis problems*, 2nd Edition, Packt Publishing, Birmingham.
10. Hastie, T; Tibshirani, R; Friedman, J (2009). *The elements of statistical learning. Data mining, inference, and prediction.* Second edition, Springer Series in Statistics. Springer, New York, 2009.
20. Apolloni, B; Bassis, S (2018). The randomness of the inferred parameters. A machine learning framework for computing confidence regions. *Inform. Sci.* 453 (2018), 239–262.
21. Martínez, AM.; Webb, GI.; Chen, S; Zaidi, NA (2016). Scalable learning of Bayesian network classifiers. *J. Mach. Learn. Res.* 17 (2016), Paper No. 44, 35 pp.
11. Bandyopadhyay, S; Pal, SK (2007). *Classification and learning using genetic algorithms. Applications in bioinformatics and web intelligence.* Natural Computing Series. Springer, Berlin, 2007.
12. Jordan, MI, Mitchell, TM (2015). Machine learning: Trends, perspectives, and prospects *Science* 17 Jul 2015: Vol. 349, Issue 6245, pp. 255-260
13. Skansi, S (2018). Introduction to deep learning. From logical calculus to artificial intelligence. Undergraduate Topics in Computer Science. Springer, Cham, 2018.
14. Goodfellow, I; Bengio, Y; Courville, A (2016). *Deep learning. Adaptive Computation and Machine Learning.* MIT Press, Cambridge, Massachusetts.
15. Chen XW, Lin X (2014). Big Data Deep Learning: Challenges and Perspectives, *IEEE Access, Volume 2, pp: 514-525.*
16. Miller DD, Brown EW (2018). Artificial Intelligence in Medical Practice: The Question to the Answer? *Am J Med.* 2018 Feb;131(2):129-133.
17. Lantz, B. (2020). Building Better Machine Learning Models (chapter in-preparation), *Principles and Methods for Data Science, Handbook of Statistics*, Volume 42, (Eds. Arni S.R. Srinivasa Rao and C.R. Rao), Elsevier-North Holland, Amsterdam
18. Van Messem A (2020). Support Vector Machines, a robust prediction method with applications in bioinformatics (in-preparation), *Principles and Methods for Data Science, Handbook of Statistics*, Volume 42, Elsevier-North Holland, Amsterdam (Eds. Arni S.R. Srinivasa Rao and C.R. Rao)
19. Abarbanel, HDI.; Rozdeba, PJ.; Shirman, S (2018). Machine learning: deepest learning as statistical data assimilation problems. *Neural Comput.* 30 (2018), no. 8, 2025–2055.
22. Nielsen, F (2018). What is... an information projection? *Notices Amer. Math. Soc.* 65 , no. 3, 321–324.
23. Bergstra, J; Bengio, Y (2012). Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* 13 (2012), 281–305
24. Bouveyron, C; Latouche, P; Mattei, PA (2018). Bayesian variable selection for globally sparse probabilistic PCA. *Electron. J. Stat.* 12 (2018), no. 2, 3036–3070.

25. Veloso de Melo, V; Banzhaf, W (2018). Automatic feature engineering for regression models with machine learning: an evolutionary computation and statistics hybrid. *Inform. Sci.* 430/431 (2018), 287–313.
26. Vidyasagar, M (2014). Machine learning methods in the computational biology of cancer. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 470 (2014), no. 2167, 20140081, 25 pp.
27. Athey, S; Imbens, G (2016). Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* 113 (2016), no. 27, 7353–7360.
28. Lee, J.; Wu, Y.; Kim, H (2015). Unbalanced data classification using support vector machines with active learning on scleroderma lung disease patterns. *J. Appl. Stat.* 42 (2015), no. 3, 676–689.
29. Kalidas, Y (2020). Data Science - Concepts, Algorithms and Practice (chapter in preparation), *Principles and Methods for Data Science, Handbook of Statistics*, Volume 42, (Eds. Arni S.R. Srinivasa Rao and C.R. Rao), Elsevier-North Holland, Amsterdam.
30. Govindaraju V and C. R. Rao (Editors) (2013). Machine learning: theory and applications. *Handbook of Statistics*, 31. Elsevier/North-Holland, Amsterdam, 2013. xxiv+525 pp.
31. Bishop, CM (2013). Model-based machine learning. *Philos. Trans. R. Soc. Lond. Ser. A*
38. Cherkassky, M (2009). Application of machine learning methods to medical diagnosis. *Chance* 22 (2009), no. 1, 42–50.
39. Varun LK; Ryan S; David E (2018). Holographic diagnosis of lymphoma, *Nature Biomedical Engineering*, 2: 631–632 (2018)
40. Murthy, KR; Singh, S; Tuck, D; Varadan, V (2019). Bayesian Item Response Theory for Cancer Biomarker Discovery, *Integrated Population Biology and Modeling, Handbook of Statistics*, Volume 40, (Eds. Arni S.R. Srinivasa Rao and C.R. Rao), Elsevier-North Holland, Amsterdam.
41. Kurmukov, A; Dodonova, Y; Zhukov, LE (2017). Machine learning application to human brain network studies: a kernel approach. *Models, algorithms, and technologies for network analysis*, 229–249, Springer Proc. Math. Stat., 197, Springer, Cham, 2017.
- Math. Phys. Eng. Sci.* 371 (2013), no. 1984, 20120222, 17 pp.
32. Freno, BA.; Carlberg, KT (2019). Machine-learning error models for approximate solutions to parameterized systems of nonlinear equations. *Comput. Methods Appl. Mech. Engrg.* 348 (2019), 250–296.
33. LeCun Y, Bengio, Y, Hinton G (2015). Deep learning, *Nature*, 521: 436–444.
34. Erhan, D; Bengio, Y; Courville, A; Manzagol, PA; Vincent, P; Bengio, S (2010). Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.* 11 (2010), 625–660.
35. Bengio, Y; Lamblin, P; Popovici, D; Larochelle, H (2007). Greedy layer-wise training of deep networks. *Advances in Neural Information Processing Systems* 19 (NIPS'06), (Eds. Bernhard Schölkopf, John Platt, and Thomas Hoffman). pages 153–160.
36. Yaron G, Yair H, Omri B, Guy N, Nicole F, Dekel G, Lina BS, Peter MK, Susanne BK, Martin Z, Lynne MB; Karen WG (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, 25: 60–64 (2019).
37. Komorowski, M; Leo AC; Badawi, O; Gordon, AC; Faisal, AA (2018). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, 24: 1716–1720 (2018)
42. Saha, A; Dewangan, C; Narasimhan, H; Sampath, S; Agarwal, S (2014). Learning score systems for patient mortality prediction in intensive care units via orthogonal matching pursuit. *Proceedings of the 13th International Conference on Machine Learning and Applications (ICMLA)*, 2014, pp:93-98.
43. He, JY; Wu, X; Jiang, YG; Peng, Q; Jain, R (2018). Hookworm detection in wireless capsule endoscopy images with deep learning. *IEEE Trans. Image Process.* 27 (2018), no. 5, 2379–2392.
44. Gurve, D; Krishnan, S (2018). Deep learning of EEG time-frequency representations for identifying eye states. *Adv. Data Sci. Adapt. Anal.* 10 (2018), no. 2, 1840006, 13 pp.
45. Carneiro, G; Nascimento, JC.; Freitas, A (2012). The segmentation of the left ventricle of the heart from ultrasound data

- using deep learning architectures and derivative-based search methods. *IEEE Trans. Image Process.* 21 (2012), no. 3, 968–982.
46. Rueda, A; Krishnan, S (2018). Clustering Parkinson's and age-related voice impairment signal features for unsupervised learning. *Adv. Data Sci. Adapt. Anal.* 10 (2018), no. 2, 1840007, 24 pp.
47. Ustun, B; Rudin, C (2016). Supersparse linear integer models for optimized medical scoring systems. *Mach. Learn.* 102 (2016), no. 3, 349–391.
48. Agarwal, S; Niyogi, P (2009). Generalization bounds for ranking algorithms via algorithmic stability. *Journal of Machine Learning Research*, 10:441-474, 2009.
49. Tu, C (2019). Comparison of various machine learning algorithms for estimating generalized propensity score. *J. Stat. Comput. Simul.* 89 (2019), no. 4, 708–719.
50. Patel, H; Thakkar, A; Pandya, M; Makwana, K (2018). Neural network with deep learning architectures. *J. Inf. Optim. Sci.* 39 (2018), no. 1, 31–38.
51. Polson, NG.; Sokolov, V (2017). Deep learning: a Bayesian perspective. *Bayesian Anal.* 12 (2017), no. 4, 1275–1304.
52. Jiequn, H, Arnulf, J, Weinan, E (2018). Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. USA* 115 (2018), no. 34, 8505–8510.
53. Agarwal, N; Bullins, B; Hazan, E (2017). Second-order stochastic optimization for machine learning in linear time. *J. Mach. Learn. Res.* 18 (2017), Paper No. 116, 40 pp.
54. Sirignano, J, Spiliopoulos, K (2018) DGM: a deep learning algorithm for solving partial differential equations. *J. Comput. Phys.* 375 (2018), 1339–1364.
55. Ye, J., Chul, Han, Yoseob; Cha, E (2018) Deep convolutional framelets: a general deep learning framework for inverse problems. *SIAM J. Imaging Sci.* 11 (2018), no. 2, 991–1048.
56. Pan, S; Duraisamy, K (2018). Data-driven discovery of closure models. *SIAM J. Appl. Dyn. Syst.* 17 (2018), no. 4, 2381–2413.
57. Freno, BA.; Carlberg, KT (2019). Machine-learning error models for approximate solutions to parameterized systems of nonlinear equations. *Comput. Methods Appl. Mech. Engrg.* 348 (2019), 250–296.
58. Mahsereci, M; Hennig, P (2017). Probabilistic line searches for stochastic optimization. *J. Mach. Learn. Res.* 18 (2017), Paper No. 119, 59 pp.
59. Schwab, C; Zech, J (2019) Deep learning in high dimension: neural network expression rates for generalized polynomial chaos expansions in UQ. *Anal. Appl.* (Singap.) 17 (2019), no. 1, 19–55.
60. Srivastava, N; Salakhutdinov, R (2014). Multimodal learning with deep Boltzmann machines. *J. Mach. Learn. Res.* 15 (2014), 2949–2980.
61. Salakhutdinov, R; Hinton, G (2012). An efficient learning procedure for deep Boltzmann machines. *Neural Comput.* 24 (2012), no. 8, 1967–2006.
62. Jiang, B; Wu, TY; Zheng, C; Wong, WH (2017). Learning summary statistic for approximate Bayesian computation via deep neural network. *Statist. Sinica* 27 (2017), no. 4, 1595–1618.
63. Deng, Y; Bao, F; Deng, X; Wang, R; Kong, Y; Dai, Q (2016). Deep and structured robust information theoretic learning for image analysis. *IEEE Trans. Image Process.* 25 (2016), no. 9, 4209–4221.
64. Baldi, P; Sadowski, P; Lu, Z (2018). Learning in the machine: random backpropagation and the deep learning channel. *Artificial Intelligence* 260 (2018), 1–35.
65. Chao, D, Zhu, J, Zhang, B (2018), Learning deep generative models with doubly stochastic gradient MCMC. *IEEE Trans. Neural Netw. Learn. Syst.* 29 (2018), no. 7, 3084–3096.
66. Poggio, T; Smale, S (2003). The mathematics of learning: dealing with data. *Notices Amer. Math. Soc.* 50 (2003), no. 5, 537–544.

Table 1. Data Used for Computing Probabilities to Train a Machine

Patient Characteristics				Treatment Options						Pregnancy Outcomes	
Clinical	Demographic	Genetic	Other	OI_1	OI_2	OI_3	...	OI_p	IVF	Conceived	Live birth
c_1	d_1	g_1	o_1	y/n	y/n	y/n		y/n	y/n	y/n	y/n
c_2	d_2	g_2	o_2	y/n	y/n	y/n		y/n	y/n	y/n	y/n
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_k	d_l	g_m	o_n	y/n	y/n	y/n		y/n	y/n	y/n	y/n

Note: There are k –clinical, l –demographic, m –genetic, and n –other types of patients characteristics, where the values of k , l , m , and n need not be the same. A given patient could have any one of the all possible combinations ($k \times l \times m \times n$) of patient characteristics, where $\times$ is the symbol for multiplication, for example, $\{c_5, d_4, g_1, o_1\}$, $\{c_k, d_1, g_3, o_7\}$, $\{c_1, d_3, g_6, o_2\}$ etc., There are p types of ovulation induction (OI) treatments and IVF options available and these are recorded as presence (y) or absence (n) corresponding to each of the combination of patient characteristics. Finally for each of these patient treatment combinations whether the couple conceived (y) or not conceived (n) and if conceived then whether a live birth observed (y) or not (n) is reported.