A Preeclampsia Diagnosis Approach using Bayesian Networks

Mário W. L. Moreira[1,2], Joel J. P. C. Rodrigues[1,3], Antonio M. B. Oliveira[2], Ronaldo F. Ramos[2], Kashif Saleem[4]
[1] Instituto de Telecomunicações, University of Beira Interior, Portugal
[2] Instituto Federal do Ceará, Brazil
[3] University of Fortaleza (UNIFOR), Ceará, Brazil
[4] Center of Excellence in Information Assurance (CoEIA), King Saud University (KSU), Riyadh, Kingdom of Saudi Arabia

mario.moreira@it.ubi.pt; joeljr@ieee.org; amauroboliveira@gmail.com; ronaldo@ifce.edu.br; ksaleem@ksu.edu.sa

Abstract— **Hypertension is the main cause of maternal death. Preeclampsia can affect pregnant women before or during pregnancy. Identification of patients with higher risk for preeclampsia allows some precautions that are taken to prevent its severe disease and subsequent complications. In medicine, there are different situations that deal with a large range of information, which needs a thorough assessment to be able to help experts in the decision-making process. Smart decision support systems allow grouping all existing information and finding pertinent information from it. Bayesian networks offer models that allow the information capture and handle situations of uncertainty. This paper proposes the construction of a system to support intelligent decision applied to the diagnosis of preeclampsia using Bayesian networks to help experts in the pregnant's care. The processes of qualitative and quantitative modeling to the construction of a network are also presented. The main contribution of this work includes the presentation of a Bayesian network built to help decision makers in moments of uncertainty in care of pregnant women.**

Keywords— Decision support systems; Bayesian networks; Pregnancy; Hypertencion; Modeling

I. Introduction

The quality of care provided by health services for pregnant women is the most important action to control maternal mortality. Access to these services presents a major impact on these reductions, since they have enough quality to identify risks. Information systems offer the ability to monitor and evaluate the pregnant health notifying the experts in good time about a given complication that can occur during pregnancy. This allows the physicians make better decisions on diagnosis and establish better medical procedures and treatment.

The most frequently diseases during pregnancy are infectious, especially those that reach the urinary tract. These diseases can cause severe complications like increasing the risk of miscarriage and anticipation of the birth labor. However, the main concerns of obstetricians are related to metabolic syndromes such as preeclampsia and gestational diabetes, which are more fatal for both mothers and babies.

Preeclampsia occurs when a pregnant woman has high blood pressure (above 140/90 mmHg) at any time after the 20th pregnancy week and disappears before 12 weeks postpartum [*1*]. Besides high blood pressure, other complications such as excessive protein in the urine should occur in a diagnosis of preeclampsia. In [*2*], the authors identify high-risk pregnancy complaints and propose a method that analyses the Doppler signal to identify these conditions. Conclusions show that complications with pregnancy are associated with hypertensive disorders (preeclampsia), intra-uterine growth restriction of fetus, and gestational diabetes mellitus. In [*3*], the importance of a reliable diagnosis with accurate measurement of blood pressure and proteinuria is discussed. The authors also present cases where preeclampsia goes undiagnosed due to lack of appropriate equipment and limited resources in laboratories. Cheng *et al.* [*4*] analyze the relationship between the blood pressure and risk factors of pregnant women. To determine the effects over gestational age a varying coefficient model was established. Results show that effects of known risk factors change with gestational age. This result is an important knowledge for understanding the causes of gestational hypertension. In [*5*] the authors investigate the alterations in the progress of a normal pregnancy and pregnancy disorders associated with hypertension. This research work uses the Joint Symbolic Dynamics method. Mukherjee *et al.* [*6*] use Discriminant Analysis and k-means clustering to predict preeclampsia based on lipid parameters. This technique is used to separate the pregnant women in two groups named preeclampsia and control, so that a new patient can be classified into any of these groups according to estimated values of the parameters. In this context, several efforts have been performed in order to develop a system for giving support to experts in pregnant's care. Then, this work aims to investigate how Bayesian networks can support clinicians to identify high-risk pregnancy. It proposes the use of a statistical model based on Bayesian networks to better classify the seriousness of a problem helping the decision-makers in uncertainty moments. These systems applied to healthcare offer the possibility to monitor and evaluate pregnant's health and notify any complications that can occur during the pregnancy in a due time. The main contribution of this work includes a proposal of a smart system based on Bayesian networks to support decision makers in pregnant women monitoring using the Noisy-or classifier.

The remainder of this paper is organized as follows. Section II describes the use of Bayesian networks on healthcare, while Section III shows the network modeling and the construction of the tables of probabilities. Performance comparison of the proposed method, in comparison with other available approaches, and results analysis are considered in Section IV. Finally, Section V provides the conclusion and suggestions for future works.

II. BAYESIAN NETWORKS ON HEALTHCARE

This section addresses the use Bayesian networks in healthcare. The use of this technique on the proposal presented in this paper is also considered.

Bayesian networks are a methodology for the construction of systems that rely on probabilistic knowledge. These systems function with uncertain and incomplete knowledge through of Bayesian Probability Theory. Teles *et al.* [*7*] propose the use of a context-aware platform based on Bayesian networks to support the experts' decision-making in public health systems. This study is focused on scenarios of dengue. Results show that the use of ontologies together with a Bayesian networks approach makes the prediction more refined. Bobba *et al.* [*8*] present a data based DSS that uses a Bayesian approach to merge gene expression data into prognostic models. This system integrates information from earlier experiments to predict the disease state. Future works proposes the implementation of this research on other pathologies for prevention and treatment. Huang *et al*. [*9*] use data mining techniques to extract rules and relationships between diseases using patient medical records. This study used an ensemble of classifiers like Naïve Bayes and J-48 to try to improve the prediction performance of several diseases. Results showed a small improvement of the accuracy, sensitivity, and F-measure. Hannan *et al.* [*10*] describe an intelligent DSS built with artificial intelligence mechanisms. This system uses a Bayesian inference mechanism to identify the confidence level for each possible cause. The preliminary results show that this system is able to assist people in decision-making process.

Kachroo *et al.* [*11*] compared three classifiers using machine-learning techniques to project cancer incidence and mortality. This research evaluates the performance of these classifiers, examining the accuracy of each method. The results show that Naïve Bayes classifier provides the best results. Researchers plan advances in developing nonlinear models for projection of future cancer occurrences. Qian *et al*. [*12*] developed a framework that uses the naive Bayesian classifier to measure the risk probability caused by lesions in the coronary artery. Results show a potential to reduce medical expenses avoiding unnecessary test and treatment. In [*13*], the authors propose a computational model to classify stages of heart failure. They evaluate the best classifier like Naïve Bayesian, Support Vector Machine, and Radial Basis Function Network to offer a quantitative tool to facilitate the early diagnosis. Authors proposed that further works should analyze other intelligent algorithms to keep improving the best found classifier. In [*14*], the authors use Bayesian networks to assist the diagnosis of social anxiety disorder. This approach is modeled using conditional probability tables. Results show the model can be efficient for diagnosis of anxiety disorder.

Based on the related literature analysis, next section will describe a mathematical model based on Bayesian networks that can assist decision makers in uncertain times to diagnose and evaluate the gravity of hypertension in pregnant women.

III. BAYESIAN NETWORKS MODELING OF HYPERTENSION

Information about the diagnosis of hypertension in pregnant women is essential to create a smart system designed to support the decision-making in healthcare. In this proposed model the network nodes three groups can be considered: *i*) Risk factors – variables that activate physiological mechanisms; *ii*) Physiological mechanisms – functioning model of diseases related to preeclampsia; and *iii*) Symptoms/exams – physical manifestations of disease and test results. The network nodes are presented in Table I.

TABLE I. DISTRIBUTION OF NODES THAT MAKES UP THE PROPOSED MODEL IN THREE LEVELS.

	Network nodes of model		
	Risk Factors	***Physiological mechanisms***	***Symptoms /exams***
	Family history of preeclampsia	Protein in the urine (proteinuria)	Headache
	First pregnancy	Low platelet count	Epigastric pain
	Age	Impaired liver function	Nausea/vomiting
	Multiple pregnancy	Signs of kidney problems besides the urine protein	Blurring of Vision
	Obesity	Fluid in the lungs (pulmonary edema)	Giddiness
	Hypertension	New-onset headaches	Hyperflexia
	Migraine	Visual disturbances	Edema
	Diabetes type 1 or type 2		Oliguria
	Kidney disease		Hypertencion
	Tendency to develop blood clots		Proteinuria
	Autoimmune disease		

The expert can put (or not) nodes in the network according to the amount of the available information. After realization of the inference process, the expert will obtain the conditional probability of the other network nodes. Proteinuria and Hypertension nodes will provide the biggest contribution for decision-making process. Nevertheless, the other nodes may also be useful in the evaluation of the pregnant woman.

A subsequent stage of a Bayesian network structure creation is the specification of their probabilities. These probabilities can be obtained in two ways: from specialists or from an automatic learning process from a database. It is also possible to combine the two alternatives. Marginal probabilities are the easiest way to find and correspond to nodes without parents. These probabilities correspond to the prevalence of diseases in pregnant seeking medical assistance. Graphically, these nodes are represented in the upper part of the network. Based on the research of Kumar *et al*. [*15*] tables of conditional and marginal probabilities were built in Tables II and III.

TABLE II. MARGINAL PROBABILITIES TO NODES WITHOUT PARENTS.

Table of Marginal Probabilities ($n = 164$)		
Symptom	***True***	***False***
Headache	50%	50%
Epigastric pain	6%	94%
Nausea/vomiting	18%	82%
Blurring of Vision	20%	80%
Giddiness	15%	85%
Hyperflexia	29%	71%
Edema	46%	54%
Oliguria	3%	97%
Hypertencion	86%	14%
Proteinuria	80%	20%

Many network nodes have conditional probability in relation to their parents. Diagnostic tests correspond that may or may not be observed, depending on the available equipment. The physiological nodes group represents disease characteristics that may be difficult to measure but are important variables for the modeling.

A Bayesian network is defined by its structure and the corresponding probabilistic model. It determines unequivocally the joint distribution for the variables describing as mentioned in Eq. 1.

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i | parents(x_i)) \qquad (1)$$

For certain types of nodes, the conditional probabilities can be calculated from other probabilities instead of being specified directly. The classifier Noisy-or allows such calculation. For medical problems, such representation is appropriate when there is a disease with several risk factors/causes or a symptom caused by various diseases. Using a node *D* with d (true) and $\bar{d}$ (false) representing a disease, its causes are $R_1, R_2, \dots, R_n$, the probabilities for *D* are given by the joint table conditional probabilities (Eq. 2).

$$P(D | R_1, R_2, \dots, R_n) \qquad (2)$$

The Noisy-or model allows the calculation of the joint table of conditional probabilities from probability given in (2) for each parent node R_i, whilst respecting the restriction on relatives laid earlier. Table III represents the distribution of some conditional probability of patients with preeclampsia given particular symptom occurring.

TABLE III. CONDITIONAL PROBABILITIES OF D.

Probability to develop preeclampsia presenting particular symptom ($n = 164$)		
Symptom	***True***	***False***
Headache	41%	59%
Epigastric pain	0%	100%
Nausea/vomiting	22%	78%
Blurring of Vision	12%	88%
Giddiness	20%	80%
Hyperflexia	0%	100%
Edema	65%	35%
Oliguria	0%	100%
Hypertencion	88%	12%
Proteinuria	65%	35%
Both Hypertencion and Proteinuria	65%	35%

Fig. 1 illustrates an example of the application of Noisy-or to a binary node *D*. Initial values for the calculation are conditional probabilities using (2). Last step is to calculate the joint probabilities for *D*, using for this the connection probabilities.

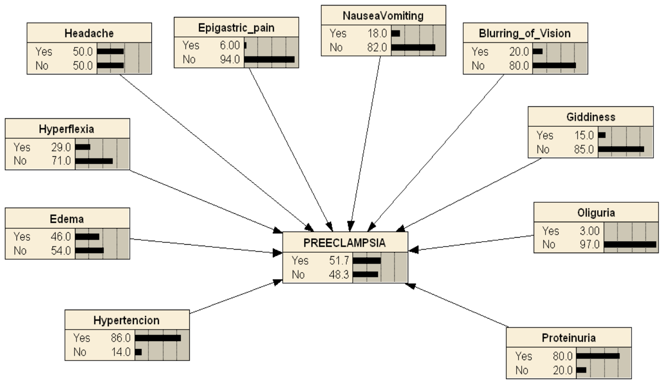

Fig. 1. *Network with the symptoms of preeclampsia.*

The use of this Bayesian network simulates various situations that may happen during the care of pregnant in a clinic. A physician inserts the pregnant symptoms on the network and analyzes the values of the conditional probabilities. At the last step, it is observed the joint probability for preeclampsia calculated by conditional probabilities. The result gives to the expert the percentage that a pregnant woman has in order to develop the disease. Figure 2 shows an example that contains the main observations and the results of the network.

IV. PERFORMANCE EVALUATION OF THE PROPOSED METHOD

For the evaluation and validation of the proposed method a case study is performed. It will use information about symptoms of patients in different hypertension severity.

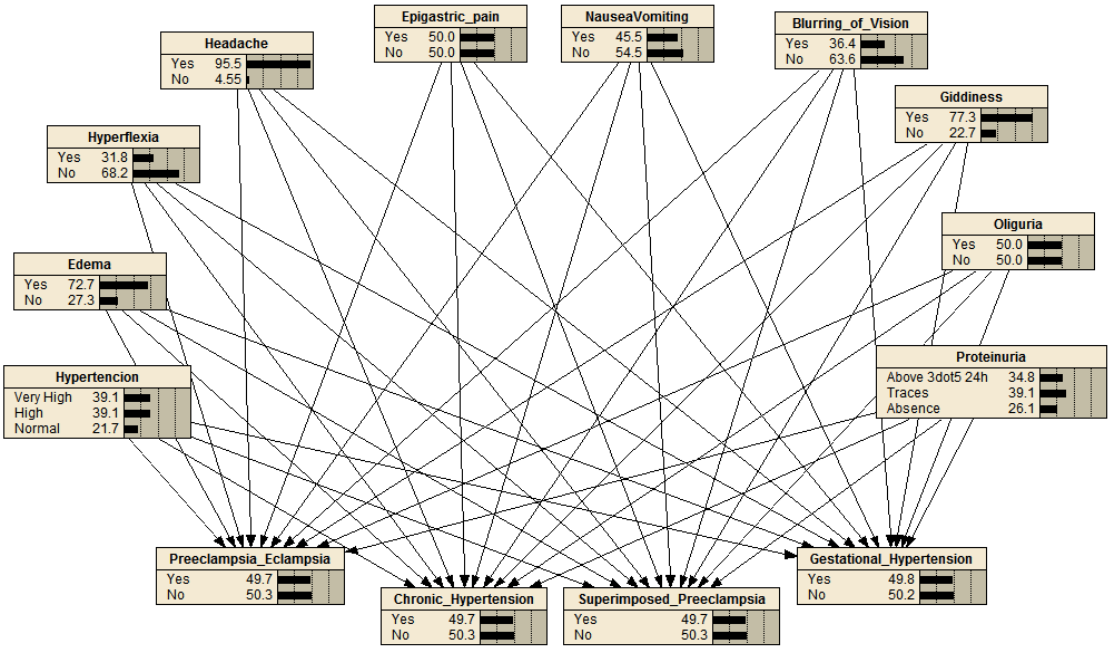

Fig. 2. *Bayesian network to support physicians in the process of decision-making of preeclampsia.*

For the study 20 pregnant women were recruited. Fig. 2 shows the final model network, which is used as a basis for inference of new cases. For evaluation of this approach the networks generation based in information about symptoms given by the health specialist were considered. Two cases sample were randomly chosen for validate the proposed system, observing the highest probability of each degree of severity in gestational hypertension.

A. Case 1: Patient with preeclampsia/eclampsia diagnostics

In this case, the pregnant woman had high blood pressure (BP > 140/90 mmHg), pulmonary edema, hyperflexia, headache, nausea or vomiting, giddiness and proteinuria (>3,5g/24h).

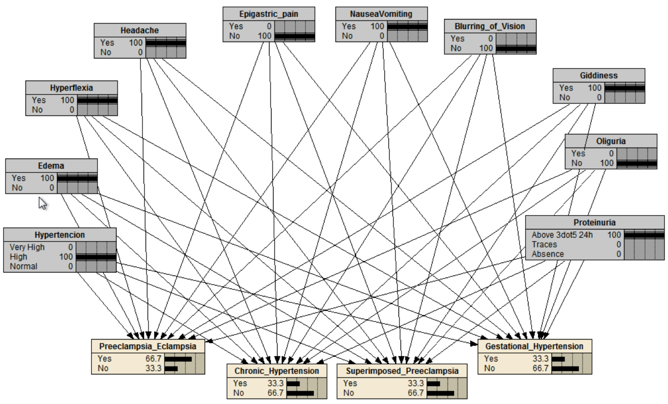

Fig. 3. *Model of network giving to the experts a probability for each severity of hypertension.*

The network presented in Fig. 3 shows a chance of 66.7% to this patient to present preeclampsia or eclampsia during pregnancy.

B. Case 2: Patient with chronic hypertension with superimposed preeclampsia.

The chronic hypertension with superimposed preeclampsia presents *i)* emergence of proteinuria (> 0,3g/24h) after the gestational age of 20 weeks in a patient with chronic hypertension; *ii)* an additional increase in proteinuria in those who have had increase previously; *iii)* a sudden increase in blood pressure in who had previously controlled levels; or *iv)* clinical or laboratory abnormality characteristic of preeclampsia. In this case the pregnant woman presented high blood pressure (BP > 140/90 mmHg), hyperflexia, headache, giddiness, oliguria, and traces of proteinuria.

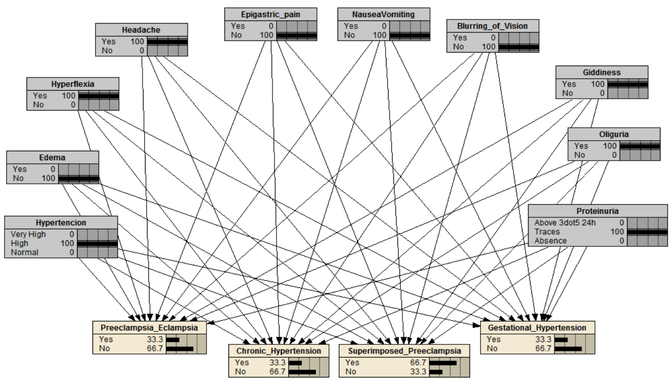

Fig 4. Model of network preseting a probability of 66,7% for chronic hypertension with superimposed preeclampsia.

Fig 4. also presents a probability of 66,7% for chronic hypertension with superimposed preeclampsia. This probability can be improved with the inclusion of new data. The presented model helps to better understand the patient's condition, assisting the diagnostic or prognostic decisions in order to reduce the uncertainty of current or future condition of the patient. This approach also requires varied sources as medical knowledge and experience. This approach requires more clinical data to be better evaluated and compared with other systems. However, with a large volume of data another type of probabilistic approach is required. Considering all these conditions, data mining can be a way to improve certainty in the moment of decision-making.

Preeclampsia is very difficult to diagnose because it can occur even without an increase of blood pressure and without the presence of protein in the urine, but research is advancing and the joint cooperation including technology, knowledge and experience of health experts is an important path to tread.

V. Conclusion and Future Work

This work focused on the construction of a smart system designed to support a medical decision for pregnant healthcare. The proposed decision support system used probabilistic concepts for decision-making. A Bayesian network for the diagnosis of preeclampsia was presented. The network structure was obtained from medical references. Noisy-or model was also considered in this work. The operation of the network shows that, in certain cases, this type of modeling can be used profitably, especially when it has a large number of parents, and when parents have characteristics in common.

Further research work will consider other Bayesian classifiers and evaluate the network using real cases and the corresponding experts evaluation. This evaluation will provide different views regarding parts of the network and will contribute for further deployments. It is also proposed to carry out practical experiments with the network, sensitivity analysis, and development of a user interface and corresponding application.

Acknowledgements

Part of this work has been partially supported by the *Instituto de Telecomunicações*, Next Generation Networks and Applications Group (NetGNA), Portugal, by National Funding from the FCT - *Fundação para a Ciência e a Tecnologia* through the UID/EEA/50008/2013 Project, by the National Plan of Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (12-INF2817-02), by Laboratório de Redes de Computadores (LAR-A), Aracati, Brazil, and the Scientific Mobility Program - *Ciências sem Fronteiras* - of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

References

[1] Y. Yoneyama, R. Sawa, S. Suzuki, D. Doi, K. Yoneyama, Y. Otsubo, and T. Araki, "Relationship between plasma malondialdehyde levels and adenosine deaminase activities in preeclampsia," *International Journal of Clinical Chemistry and Diagnostic Laboratory Medicine*, vol. 322, no. 1-2, pp. 169-173, August 2002.

[2] R. Das, P. Vajinepalli, R. Sisodia, and L. Gupta, "Automatic Doppler Signal Analysis to Assess Utero-Placental Circulation for Identifying High Risk Pregnancies," in *35th Annual International Conference of the IEEE EMBS*, Osaka, 2013, pp. 3905-3908.

[3] H. D. Tagare, K. Rood, and I. A. Buhimschi , "An Algorithm to Screen for Preeclampsia Using a Smart Phone," in *2014 Health Innovations and Point-of-Care Technologies Conference*, Seattle, 2014, pp. 52-55.

[4] W. Cheng, L. Fang, L. Yang, H. Zhao, P. Wang, and J. Yan, "Varying Coefficient Models for Analyzing the Effects of Risk Factors on Pregnant Women's Blood Pressure," in *13th International Conference on Machine Learning and Applications*, Detroit, 2014, pp. 55-60.

[5] Voss A., Seeck A., and Baumert M., "Altered interactions of heart rate and blood pressure during normal and abnormal pregnancy," in *32nd Annual International Conference of the IEEE EMBS*, Buenos Aires, 2010, pp. 1695-1698.

[6] R. Mukherjeea, C. D. Ral, C. Chakrabortya, S. Dasgupta, and K. Chaudhury, "Clinical biomarker for predicting preeclampsia in women with abnormal lipid profile: statistical pattern classification approach," in *International Conference on Systems in Medicine and Biology*, Kharagpur, 2010, pp. 397-401.

[7] G. Teles, C. Oliveira, R. Braga, L. Andrade, R. Ramos, P. Cunha, and M. Oliveira, "Using Bayesian Networks to improve the Decision-Making Process in Public Health Systems," in *EEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Natal, 2014, pp. 565-570.

[8] S. S. Bobba, A. Zollanvari, and G. Alterovitz, "Bayesian Prognostic Model for Genomic Discovery in Bipolar Disorder," in *Health Innovations and Point-of-Care Technologies Conference*, Seattle, 2014, pp. 247-250.

[9] F. Huang, S. Wang, and C. Chan, "Predicting Disease By Using Data Mining Based on Healthcare Information System," in *IEEE International Conference on Granular Computing*, Hangzhou, 2012, pp. 191-194.

[10] B. Hannan, X. Zhang, and K. Sethares, "iHANDs: Intelligent Health Advising and Decision-Support Agent," in *IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Warsaw, 2014, pp. 294-301.

[11] S. Kachroo, W. W. Melek, and C. J. Kurian, "Evaluation of Predictive Learners for Cancer Incidence and Mortality," in *The 4th IEEE International Conference on E-Health and Bioengineering*, Iași, 2013, pp. 1-6.

[12] Z. Qian, I. Marvasty, S. Rinehart, and S. Voros, "A Lesion-Specific Coronary Artery Calcium Quantification Framework for the Prediction of Cardiac Events," *EEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 5, pp. 673-680, September 2011.

[13] X. Fu, Y. Ren, G. Yang, Q. Pan, S. Gong, L. Li, J. Yan, and G. Ning, "A Computational Model for Heart Failure Stratification," in *Computing in Cardiology*, Hangzhou, 2011, pp. 385-388.

[14] Z. S. Estabragh, M. M. R. Kashani, F. J. Moghaddam, S. Sari, and K. S. Oskooyee, "Bayesian Network Model for Diagnosis of Social Anxiety Disorder," in *2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, Atlanta, 2011, pp. 639-640.

[15] S. Kumar, D. Bansal, D. Hota, M. Jain, P. Singh, and B. L. Pandey, "Assessment of clinical outcomes and prescribing behavior among inpatients with severe preeclampsia and eclampsia: An Indian experience," *Indian Journal of Pharmacology*, vol. 46, no. 1, pp. 18-23, January 2014.