International Journal Scientific and Profesional (IJ-ChiProf)
Vol 4 Issue 2 March-May 2025, pp: 487-492
ISSN: 2829-2618 (Online)
DOI: https://doi.org/10.56988/chiprof.v4i2.80

Implementation of the Bayesian Network Algorithm to Predict Chronic Diseases Using Electronic Medical Record Data at UPTD RSD Besemah, Pagar Alam City

Angga Putrawansyh. PB [1], Tata Sutabri [2]
[1,2]Magister of Informatics Engineering, Universitas Bina Darma, Indonesia
Email: ang.putra777@gmail.com[1] , tata.sutabri@gmail.com[2]

Article Info

Article history:
Received April 20, 2025
Revised April 21, 2025
Accepted April 22, 2025

Keywords:
Bayesian Network
Chronic Diseases
Electronic Medical Records
Prediction

ABSTRACT

Chronic diseases are one of the leading causes of death in Indonesia and around the world. Early detection of chronic diseases poses a significant challenge for healthcare facilities, particularly in resource-limited areas such as UPTD RSD Besemah, Pagar Alam City. This study aims to implement the Bayesian Network algorithm to predict chronic diseases based on patient's electronic medical record (EMR) data. The Bayesian Network method was chosen due to its ability to model causal relationships between variables and its robustness in handling incomplete data. The data used in this research consists of secondary data from patient medical records, with attributes including age, gender, medical history, laboratory results, and lifestyle factors. The research involves data collection, preprocessing, Bayesian network structure formation, and model Performance evaluation using accuracy, precision, and recall metrics. The results indicate that the Bayesian Network model can accurately predict chronic diseases such as diabetes mellitus, hypertension, and heart disease. Implementing this predictive system is expected to assist medical personnel in clinical decision-making and enhance the effectiveness of preventive healthcare services.

Corresponding Author:
Tata Sutabri
Department of Informatics Engineering,
Bina Darma University,
Jalan A.Yani , Kota Palembang 3551
Email: tata.sutabri@gmail.com

1. INTRODUCTION

Chronic diseases such as diabetes mellitus, hypertension, and cardiovascular disease are among the leading causes of morbidity and mortality worldwide [1]. In Indonesia, the prevalence of these diseases continues to rise, placing a significant burden on the national healthcare system[2]. Early detection and risk prediction of chronic diseases are key components in effective prevention and management strategies [3].

The advancement of health information technology has enabled the collection and storage of patient data in the form of Electronic Medical Records (EMRs) [4]. These records contain rich clinical information that can be utilized for predictive analytics to support better medical decision-making [5]. Integrating this data with artificial intelligence techniques such as machine learning allows healthcare systems to become more responsive and personalized [6].

The Bayesian Network (BN) is a probabilistic graphical model capable of representing causal relationships between variables and handling uncertainty in medical data [7]. BN has been applied in various studies to predict chronic diseases, such as coronary heart disease, with promising results [8]. This model can integrate medical knowledge and historical data to estimate individual disease risk [9] The success of electronic medical record implementation lies in understanding the functional components of information systems as outlined by Sutabri, including input, process, output, storage, and control mechanisms [10].

International Journal Scientific and Profesional (IJ-ChiProf)
Vol 4 Issue 2 2025, pp: 487-492
ISSN: 2829-2618 (Online)

Implementing BN in a local context, such as at UPTD RSD Besemah in Pagar Alam City, can provide more specific insights into regional disease patterns [11]. However, challenges such as data quality, resource limitations, and the need for local model validation must be carefully addressed [12].

This study aims to develop and implement a BN model to predict the risk of chronic diseases based on EMR data at UPTD RSD Besemah [13]. The model is expected to serve as a decision support tool in clinical practice and assist in the planning of more targeted public health interventions[14]. Literature reviews also indicate that probabilistic approaches are more effective in dealing with complex and incomplete data scenarios[15]. Tata Sutabri emphasized that information systems in healthcare must not only be accurate and timely but must also align with user requirements to ensure their sustainability and adoption by health institutions [16]. In his research, Sutabri illustrated that the core value of health information systems is to support decision-making at every level of healthcare service, from diagnosis to long-term treatment management [17]

2. RESEARCH METHOD

2.1. Research Approach

This study employs a mixed-methods approach, integrating both quantitative and qualitative methodologies. The quantitative approach uses the Bayesian Network algorithm to analyse electronic medical record (EMR) data. Meanwhile, the qualitative approach seeks insights from healthcare professionals regarding chronic disease prediction in daily clinical practice.

2.2. Research Location and Period

The research was conducted at UPTD RSD Besemah, Pagar Alam City, from January to March 2025. Interviews and direct observations were conducted on-site with hospital staff, while the data modelling process was performed separately using analytical software tools.

2.3. Data Sources and Techniques

a. Quantitative Data:
Sourced from patients' electronic medical records, including medical history, diagnoses, treatments, and laboratory test results.
b. Qualitative Data:
Collected through semi-structured interviews with healthcare professionals (doctors, nurses, and hospital IT staff) regarding EMR practices, the need for disease prediction, and clinical challenges in managing chronic illnesses.

2.4. Research Stages

a. Data Collection
   1) EMR data extraction from the hospital information system.
   2) Data cleaning and categorisation based on chronic disease types.
   3) Interviews with 5–7 key informants from medical staff.
b. Data Preprocessing and Transformation
   1) Data normalisation and categorisation (e.g., numerical → categorical).
   2) Handling of missing values and data duplication.
   3) Feature selection to determine the most relevant variables for prediction.
c. Bayesian Network Model Development
   1) Implementation of algorithms using Python libraries such as pumps or pomegranate.
   2) Graph structure construction among variables.
   3) Model training and testing using hold-out or k-fold cross-validation techniques.
d. Model Evaluation
   1) The evaluation metrics used were accuracy, precision, recall, F1-score, and AUC.
   2) Interpretation of model results to assess the probability of chronic disease occurrence.
e. Qualitative Analysis
   1) Coding of interview transcripts to identify patterns, needs, perceptions, and recommendations from medical personnel.
   2) Qualitative findings are integrated with quantitative results to formulate the conclusions.

International Journal Scientific and Profesional (IJ-ChiProf)
Vol 4 Issue 2 March-May 2025, pp: 487-492
ISSN: 2829-2618 (Online)
DOI: https://doi.org/10.56988/chiprof.v4i2.80

3. RESULTS AND DISCUSSION

3.1. Result

3.1.1. Dataset Description

The dataset used in this study was derived from electronic medical records (EMRs) of patients at UPTD RSD Besemah, Pagar Alam City. It includes data from 500 patients with ten primary attributes: age, gender, family history, systolic blood pressure, diastolic blood pressure, fasting blood glucose level, and smoking status. After data cleaning, normalisation, and handling of missing values, the dataset was classified into two target categories: Chronic Disease (including diabetes, hypertension, and heart disease) and Nonchronic.

3.1.2. Bayesian Network Model Construction

This study's Bayesian Network model construction followed three main stages: structure learning, parameter learning, and probabilistic inference. These processes were implemented using the Python programming language and the pgmpy (Python Library for Probabilistic Graphical Models) library.

a. Structure Learning

This stage's goal was to define the network architecture, with nodes representing variables and edges indicating probabilistic dependencies between them.

1) Method Used: The Hill Climb Search (HCS) algorithm was chosen because it can find optimal structures through iterative search.
2) Scoring Function: Each structure was evaluated using a Bayesian Information Criterion (BIC), which balanced predictive accuracy with model complexity.
3) Technical Process:
   a) The dataset was input as pandas.DataFrame.
   b) The initial structure was a blank graph.
   c) HCS evaluated all possible edge additions, deletions, or reversals.
   d) The process terminated when there was no significant improvement in the BIC score.

b. Parameter Learning

Once the structure was formed, the next step was to compute each node's Conditional Probability Table (CPT).

1) Method Used: Maximum Likelihood Estimation (MLE) was employed to calculate the conditional probabilities of variables dependent on one or more parent nodes.
2) Application
   Data was discretised into categories (e.g., blood glucose → low, normal, high). CPTs were generated for each variable based on its parent nodes in the graph.

c. Probabilistic Inference

This stage aimed to compute the likelihood of chronic disease diagnosis based on combinations of patient attributes. The Variable Elimination algorithm efficiently calculates marginal and conditional probabilities in complex networks. The resulting probabilities provided insights into whether a patient was at high risk for a specific disease and enabled the system to issue clinical warnings or recommendations.

d. Model Visualisation Results

The model structure reveals causal relationships between variables, such as:

1) Blood_Glucose → Diabetes
2) Age → Hypertension
3) Family_History → Chronic_Disease
4) BMI → Heart_Disease

This structure illustrates how a combination of risk factors can trigger specific chronic disease diagnoses.

e. Model Evaluation

The model was evaluated using 10-fold cross-validation. The evaluation results are summarised in the following table:

Vol 4 Issue 2 2025, pp: 487-492
ISSN: 2829-2618 (Online)

Table 1. Model Evaluation

Evaluation Metric	Value (%)
Accuracy	84.7
Precision	81.3
Recall	86.9
F1-Score	84.0
AUC-ROC	0.88

An accuracy of 84.7% indicates that most model predictions were accurate. A recall score of 86.9% demonstrates the model's high sensitivity in detecting chronic disease cases (i.e., minimal false negatives). A balanced F1 score confirms the model's stability and reliability.

f. Bayesian Network Interpretation

The network structure provides logical insights into the diagnostic process:

1) High blood glucose levels yield a 92% probability of a diabetes diagnosis.
2) A combination of age over 50, high blood pressure, and family history gives a 73% probability of hypertension risk.
3) High BMI and active smoking increase the probability of heart disease by up to 68%.

These interpretations are valuable for medical professionals in making data-driven decisions beyond relying solely on intuition or experience.

3.2. Discussion

3.2.1. Consistency with Previous Studies

These findings are consistent with the study conducted by Abedi et al. (2020), which achieved similar accuracy in heart disease prediction using the Bayesian Network approach. This indicates the method's generalizability and suitability for implementation in various healthcare institutions, including UPTD RSD Besemah.

3.2.2. Model Advantages

a. Capable of handling incomplete data.
b. Able to explain causal relationships between symptoms and diseases.
c. Supports transparency in clinical decision-making.

3.2.3. Research Limitations

The dataset was limited to 500 patients, which may not fully represent the general population. Some important variables (e.g., genetic data, physical activity levels) were unavailable in the EMR system.

3.2.4. Practical Implications

This model can serve as the foundation for developing a clinical decision support system (CDSS) that assists doctors in:

a. Conducting early disease screening.
b. Providing data-driven treatment suggestions.
c. Improving diagnostic efficiency and reducing time.

4. CONCLUSION

Based on the results of the conducted research, several conclusions can be drawn as follows:

4.1. The Bayesian Network model predicts chronic diseases using electronic medical record (EMR) data. Using conditional probabilities between variables, the model accurately and transparently represents complex relationships between risk factors and disease diagnoses.

4.2. The model achieved an accuracy of 84.7%, with a recall of 86.9% and an F1-score of 84.0%. These results indicate that the developed model performs well in identifying patients at risk of chronic diseases such as diabetes, hypertension, and heart disease.
4.3. Key variables significantly contributing to chronic disease prediction include age, blood glucose level, blood pressure, body mass index (BMI), smoking status, and family history. Based on the results of structure learning, these variables have demonstrated a direct influence on diagnosis outcomes.
4.4. Implementing the Bayesian Network model provides predictions and assists medical professionals in understanding the causal relationships between variables. Thus, the model has the potential to become part of a Clinical Decision Support System (CDSS), enhancing both diagnostic accuracy and efficiency in hospital settings.
4.5. This study also demonstrates that electronic medical record data at UPTD RSD Besemah Kota Pagar Alam holds significant potential for intelligent system development, although there are still limitations regarding data volume and completeness.

5. CONCLUSION

Based on the findings and results of this study, the following recommendations are proposed for further development and practical implementation of the model:

a. Improving the Quality and Completeness of EMR particularly UPTD RSD Besemah Kota Pagar Alam, are encouraged to standardise and enrich their electronic medical records. Important variables such as physical activity, dietary patterns, and medical treatment history should be included to enhance the accuracy and contextual relevance of the prediction model.
b. Integration of the Model into Clinical Systems with Bayesian Network model can be integrated into Hospital Information Systems (HIS) or medical decision-support platforms, allowing doctors or healthcare providers to use it in real-time during initial screening or diagnostic decision-making processes.
c. Trials should be conducted using data from other hospitals or primary healthcare centres (puskesmas) with different population characteristics to assess the reliability and generalizability of the model. This will ensure that the model is applicable beyond a single institution.
d. Future research may focus on developing a user-friendly application interface so that non-technical medical personnel can easily utilise the predictive system without needing to understand the technical aspects of the Bayesian Network model.
e. Further studies are encouraged to explore hybrid approaches, such as combining Bayesian Networks with other classification algorithms like Random Forest or Gradient Boosting, to improve predictive Performance and model interpretability.

REFERENCES

[1] World Health Organization., “Noncommunicable diseases fact sheet.”
[2] K. RI., “Profil Kesehatan Indonesia 2022.”
[3] et al. Mendis, S., “Prevention and control of cardiovascular disease: An update. WHO Bulletin,” 2011.
[4] et al. Jaspers, M. W. M., “Electronic health record implementation in hospitals: A literature review. BMC Medical Informatics and Decision Making,” 2011.
[5] L. T. Nguyen, L., Bellucci, E., & Nguyen, “Electronic health records implementation: An evaluation of information system impact and contingency factors. International Journal of Medical Informatics, 83,” 2014.
[6] et al. Rajkomar, A., “Machine learning in medicine. New England Journal of Medicine, 380,” 2019.
[7] J. Pearl, “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann.,” 2020.
[8] et al. Suo, Y., “Development of a Bayesian network model for predicting coronary heart disease. Journal of the American Heart Association.,” 2024.
[9] et al. Muñoz-Valencia, C. S., “Employing Bayesian networks for disease diagnosis and prognosis: A comprehensive review. arXiv:2304.06400.,” 2023.
[10] T. Sutabri, “Struktur Sistem Informasi Rekam Medis Elektronik dalam Lingkungan Rumah Sakit Digital. Jurnal Sistem Informasi Kesehatan, 5(2), 134–141.,” 2013.
[11] et al. Lin, S., “Chronic disease risk prediction with deep reinforcement learning. Complex & Intelligent Systems, 11.,” 2025.
[12] et al. Faruqui, S. H. A., “Structure learning in Bayesian networks for chronic conditions. arXiv:2007.15847.,”

2020.
[13] et al Friedman, N., "Using Bayesian networks to analyze expression data. Journal of Computational Biology, 7(3-4), 601–620.," 2017.
[14] T. Sutabri, "Analisis Sistem Keamanan Informasi Elektronik Rekam Medis Rumah Sakit. Yogyakarta: Andi.," *2012*.
[15] et al Rajkomar, A., "Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.," 2022.
[16] T. Sutabri, "Penerapan Sistem Informasi Manajemen Rumah Sakit Berbasis Komputer. Jurnal Teknologi Informasi dan Ilmu Komputer, 2(3), 211–218.," 2014.
[17] T. Sutabri, "Peran Sistem Informasi Manajemen dalam Pengambilan Keputusan Medis. Jurnal Informasi dan Kesehatan, 3(1), 89–96.," 2015.